THE UNIVERSITY of EDINBURGH

Title	Hierarchical organisation in serial search tasks by Cebus Apella monkeys
Author	Dickinson, Anthony R.
Qualification	PhD
Year	1997

Hierarchical Organisation in Serial Search Tasks by *Cebus apella* Monkeys

Anthony R. Dickinson

Ph. D.
University of Edinburgh
1997

Abstract

The thesis reports the first demonstration of two-level hierarchical and linear organisation in six monkeys (*Cebus apella*). In behaviours elaborated over a four-year period, the hypothesis tested (McGonigle and Chalmers, 1992) was that progressive increases in task difficulty would be compensated by data reducing, economic, organisational structures. Novel touch-screen based procedures required the seriation and search of each item in a test set which, when increased in size, lead to geometrical increases in difficulty. Whilst on one scenario the subject should begin to fail as the task increases in difficulty, the performance of all subjects has shown progressive adaptation to such task requirements. This indicates an underlying dynamic process consistent with the operating hypothesis that cognitive organisation, both linear and hierarchical, are emergent responses motivated by a need for cognitive economy.

Declaration

I hereby declare that this thesis has been composed by myself and that the work herein reported was conducted as part of the research group of Dr. Brendan McGonigle of the Laboratory for Cognitive Neuroscience and Intelligent Systems at the University of Edinburgh.

Anthony R. Dickinson
Edinburgh
June 13th, 1998.

Acknowledgements

To the many who have inspired my setting out upon this 'voyage of discovery', I will be forever grateful for both their living and recorded examples. With apologies to so many more colleagues and friends who have shared their ideas and influence, I wish to name a few of my fellow researchers who are more than a little responsible for this work coming to fruition. Firstly my thanks are never given often enough to Dr. Brendan McGonigle for his providing the vision, as well as the supervision throughout this journey, and without whom none of this work would ever have been possible. Many thanks are also due to other colleagues and staff of the Laboratory for Cognitive Neuroscience and Intelligent Systems, both past and present: to Dr. Carlo DeLillo for allowing me to watch the initial struggles from a distance prior to my working with the monkeys; to Dr. Benjamin St. Johnston (and latterly Lawence Warnett) for endless help with programmimg and matters computational; to our animal technicians, Steven Brady and Della Purves for their dedication and continued devotion to both managing the day to day husbandry and later experimental supervision of the monkeys and pigeons in the laboratory; to Dr. Maggie McGonigle I offer very many thanks for advice and help along the way, including those for the 'ham' that is still to come. I would also like to thank the Department of Psychology for help in funding the latter period of my Doctoral study in the form of a salaried lectureship. Many thanks are especially due to my wife, Wai Yi (Diane), for her unending support and encouragement during the many days and nights of study and toil, whilst awaiting the successful seriation of the chapters to be found below.

The space of possible theses was indeed very large, the pathway chosen, was just but one.

Tony Dickinson
Edinburgh
June 13th, 1998.

Contents

Chapter 1

The evolution of intelligent behaviour: early mentalism versus behaviourism

Although the conclusions of Descartes were to deny the existence of a mental life to non-human animals in the early 1600s, it was not until Hume, some 150 years later, that a view of continuity between the mental processes of animals and man was offered. However, no accounts for its development were provided at this time. Indeed, serious academic debate concerning the evolution of intelligent behaviour in human and non-human animals, the nature of intelligence, and the degree to which a given species may be able to demonstrate such behaviours has a relatively short history of about 100 years. Reading the extensive literature written throughout this latter period, two recurrent themes emerge, and the various approaches to the study of the 'nature of intelligence' have been, to greater or lesser degrees influenced by them both. These are, firstly, a given researcher's position concerning the issue of the evolutionary continuity/discontinuity between species-specific behaviours; and secondly, the degree to which they were willing to accept qualitative observational and/or quantitative behavioural data for the purposes of providing evidence for their claims.

This chapter will address a brief history of ideas concerning the evolution of intelligent behaviour from the viewpoint of continuity, the theoretical precondition for a comparative psychology of cognition being provided principally by the writings of Charles Darwin and Herbert Spencer in the late 19th Century. Although a detailed discussion of the developments of this new field of comparative enquiry has been provided by Boakes (1984), it is pertinent to briefly review them here in order that the following period of experimental behaviourism and its discontinuity position become more clearly understood. The chapter concludes with a discussion of the more recent work in animal intelligence, which has become once more grounded in its evolutionary, biological, adaptation-based roots

The evolutionary stance

The idea of studying the behaviour of animals in order to better understand the human mind was independently developed and propounded

by both Darwin and Spencer in the latter half of the last century. Of the two, Darwin was by far the most influential. In developing his theory in which the primary cause of evolution was the mechanistic process of natural selection, in *The Origin of Species* Darwin (1859) hinted at his belief that the origin of the human species involved no exceptional process of special creation. Twelve years later, the reasons for this view were further elucidated with the publication of *The Descent of Man* in which Darwin (1871) was to make a stronger case for mental continuity between man and other animals. This was perhaps the first time that emphases were being placed upon the potential importance of both morphological data and species differences in learning behaviour. To this end, Darwin argued for two secondary mechanisms: one of sexual selection, the other of a principle invoking the use or disuse by an individual of its organs (including the brain) and their effects upon inheritance by its offspring. This evocation of a Lamarkian factor for the treatment of behaviour, offered both Spencer and Darwin some explanation for the perceived resemblances between an animal's instincts and the learned actions that had become habitual during its own life time. Although the task of distinguishing habit from instinct remained of great importance until the end of the century, no empirical studies were performed by either of them. This was not, however, due to any lack of theoretical consideration. In the first edition of Spencer's (1855) *Principles of Psychology* the principle underlying the transition from reflex to instinct was held to be a process of learning based upon both contiguity and frequency. A second process of learning, later known as the Spencer-Bain principle proposed that spontaneous actions which accidentally, but immediately, followed brain state changes correlated with pleasure were likely to recur. However, at this time, the statement of these ideas neither stimulated close attention, nor did they provoke any empirical investigation. The views of Darwin (and Spencer) concerning the evolution of intelligence, were nonetheless to provide the basis for the subsequent development of what was to become known as comparative psychology. For the remaining years of the 19th Century, a movement of ideas from Darwin to Romanes and later from Romanes to Morgan, provided a continuous, but evolving tradition for the study of intelligent behaviour. As Boakes (1984) has pointed out, some themes remained unchanged, for example, the aim of relating an understanding of the mind to general theories of evolution. Another, less obvious, was the continued

emphasis upon the behaviour of the individual animal and those actions which in human terms, could be called intelligent. Alternatively, emphasis might have shifted towards examining the factors involved in the development of social behaviour, as more recently expressed by several authors (Whiten and Byrne, 1988; Byrne and Whiten 1988; Dunbar, 1997), or to issues that were later labelled under the study of motivation (e.g., Stamp-Dawkins, 1992), but they did not. Other aspects, however, did undergo some development, culminating in the publication of *Animal Intelligence* by George Romanes in 1882.

As had Darwin before him, Romanes took seriously the animal behaviour data by now available, despite it remaining largely anecdotal in nature. The science of animal behaviour thus remained a form of natural history (similar to the anatomy and geology of the time) but on the question of non-human animals possessing a mind, the only objective evidence Romanes could envisage was the extent to which an animal might be said to demonstrate a capacity for choice. Romanes suggested that such an expression of choice could have been said to have taken place in an animal when its behaviour was observed to have been influenced by events within its own past experience. Although an important and oft neglected issue for many recent experimental learning paradigms, mentation was at that time still only inferred from behavioural observation alone. This continued to be true despite Romanes' proposing the criterion that an animal's 'demonstration of the ability to learn' be a prerequisite for its inclusion as a suitable candidate for the possession of intelligence. In support of the view that continuity of such an ability would be likely, and in an attempt to develop a theory of the evolution of mind, Romanes stated that:

> The lower down we go in the animal kingdom, the less capacity we find for changing adjustive movements in correspondence with changed conditions.'
>
> Romanes (1882: p.8)

In this and a later work, '*Mental Evolution in Animals*', Romanes (1884) extended his ideas to include the factors of phylogeny and ontogeny, comparing child development with that of non-human animals, but still only doing so having acquired his data via introspection and attribution respectively. He further claimed that increasingly complex behavioural

repertoires were a function of qualitatively different held competences, and also sought to differentiate sensation (reflecting tightly-coupled reflexive behaviours) from perception. From this view, it was from the seat of *relational* property evaluation that ideational thought was derived, a notion to which I will return in the next chapter. Finally, it was again Romanes who put forward the idea that mental evolution should be seen as the progressive construction of a cognitive architecture formed by/from different processes, whose interaction gave rise to the most complex forms of behaviour. Alas, despite these noble claims, neither experimental nor empirical research was provoked by Romanes work, and his evidence remained driven by the evaluation and manipulation of anecdotal data sets. Nonetheless, on the issue of continuity for the development of mental processes in human and non-human animals, Romanes had this to say concerning our (by then) presumed simian ancestors:

> 'Our knowledge of the psychology of the anthropoid apes is less than our knowledge of the psychology of any other animal. But notwithstanding the scarcity of enough material which I have to present, there is enough to show that..... in their psychology, as in their anatomy, these animals approach most nearly to Homo Sapiens.'
>
> Romanes (1882: p.245)

Paradigmatic implementation

There was, however, to be no immediate sequel to Romanes' work, and it was not until the early part of the next century that more systematic attempts were starting to be made with testing the intelligence of apes. Indeed, by this time the effects of Thorndyke's (1898) monkey work were being felt, and were promoting a supportive return to the claim of Descartes some two and a half centuries beforehand. In the face of this seemingly negative evidence there remained nonetheless, the persistent belief that some animals were more intelligent than others (e.g., Hobhouse, 1901). Critical of Thorndyke's methods and interpretations, the young Wolfgang Köhler (1925) required that an animal's test situation be completely different from anything experienced before, whilst remaining simple enough for the task to be understood. He regarded the Thorndykian puzzle-boxes and escapology problems to be too complex to allow their mechanical comprehension, but, more importantly, as with many more recent learning experiments, they denied the animals an opportunity to see what the solution might be. (In a typical experiment, a cat was placed

into a box (see Figure 1.1) with a bowl of food outside. In order to reach the food, the cat had to respond in a specified way to open a door, perhaps by pulling a lever. Initially the cat would scratch and struggle in the box, and a considerable time lapsed before it responded correctly. Across successive trials, a decline in the latency to escape would be evident).

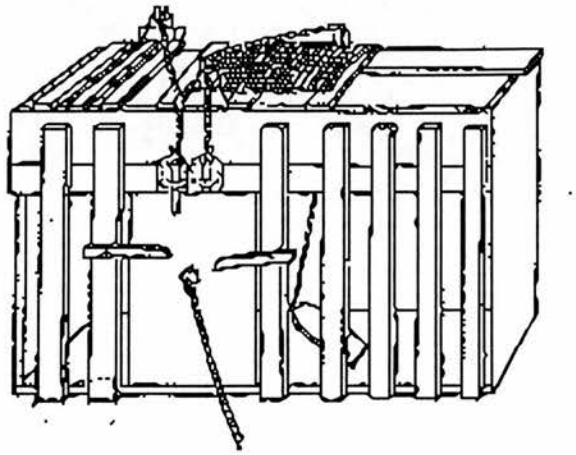


Fig. 1.1 A typical puzzle box used by Thorndyke with cats. From a sketch by Thorndyke (1898)

In his own words, for these mazes and puzzle boxes:

> 'the first time they get out is, therefore, necessarily a matter of chance... in intelligence tests of the nature of our detour (roundabout-way) experiments, everything depends upon the situation being surveyable by the subject from the outset.'
>
> Köhler (1925: p18)

Köhler was distinguishing here between intelligent, and what he called 'mechanised' behaviour, reminiscent of Spencer's and Darwin's previous allusions to 'habits' as mentioned above. In his later experiments, Köhler sought to distinguish between blind trial-and-error type learning from a more sudden and 'insightful' interpretation of behaviour. However, despite his repeated demonstrations of sequential tool-use, imitation and "insight" in chimpanzees in the laboratory (including, for example, detour-tests/box-stacking/stick-connecting) Köhler's results did not shed any light upon the issue of explaining the mental process(es) putatively involved in the successful completion of the problem-solving tasks set for his chimpanzees. At best, a new method was here being tried in an attempt to avoid the difficulties inherent in the use of Thorndyke's trial-and-error

learning experiments. This he achieved successfully, but it remains difficult to see what kind of empirical questions could be answered by such a method (particularly in the absence of 'success' for Köhler's types of task). His primary contribution was to conceive of an experimental situation in which it would be possible to allow an animal to perceive the instrumental value of a tool to be used in order to reach a goal. The important experimental feature he was to introduce was that the problem-solving environment should require some degree of forward-planning on the part of the subject in order to reach the goal. For example, he used tasks such as those illustrated in figures 1.2 - 1.3, in which the chimpanzee relocates boxes to new positions before ascending, connecting sticks prior to reaching otherwise unobtainable foodstuffs. Köhler also pointed out that the developing learning curves for trial-and-error- versus "insightful" - learning would be (and indeed were) very different; the latter showed abrupt success and stability, rather than a progressively increasing rate of success for similar tasks.

Fig. 1.2 Box stacking by chimpanzees.
(from a photogragh in Köhler, 1925)

In a later attempt to replicate these box-stacking and food reaching experiments with a novel species, Robert Yerkes worked with captive

orang-utan in the laboratory. Yerkes (1916) reported that the orang-utan would not successfully box-stack prior to seeing a demonstration (by himself), only after acquisition did he observe faster uptake (transfer) when the ape was subsequently able to use his 'newly found tools'. Once given the 'stick-test' (see Figure 1.3), in which Köhler's chimpanzees were required to connect two shorter sticks in order to reach a banana, or to pull in a longer stick with which to obtain otherwise unreachable fruit, Yerkes' orang-utan was reported to be immediately raking fruit into its enclosure. It is worthy of note, however, that the individual stages of the retrieval behaviour here might merely involve a chain of associative components leading to the attainment of such a goal, the vicarious trial-and-error interpretation being less favoured due to the complex nature of the environmental manipulation task. Indeed, having demonstrated such 'ideational behaviour' and 'instincts for mechanical ability' it was of great surprise to Yerkes that this same animal then failed to learn other supposedly 'simple' rule-learning tasks involving a multiple-choice procedure (as measured by number of trials to criterion in the pig, for example). Indeed, Yerkes was to suggest that:

> '... in this young orang-utan ideational learning tended to replace the simpler mode of problem-solving by trial-and-error. Seemingly incapable of solving his problems by the lower grade process, he strove persistently, and often vainly, to gain insight.'
>
> Yerkes (1916: p. 87)

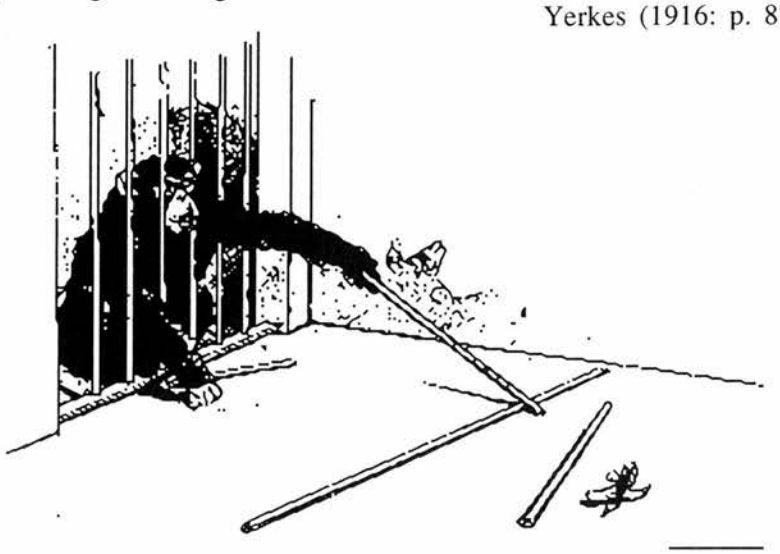


Figure 1.3. A 'stick test' problem as used by Köhler (1925). (After an illustration in Malim et. al., 1996)

Although Yerkes frequently used the term 'insight' to describe some of his apes' behaviour, he did not specify what he meant by such a term, and seemed comfortable to have produced little more than, ...

> '... the first curve of learning for an anthropoid ape..... so far as one may say by comparing it with the curve for various learning processes exhibited by other animals, it is indicative of ideation of a higher order and possibly of reasoning.'
>
> Yerkes (1916: p. 131)

Indeed, the notion that his animals had perceived objects and developed their 'appropriate relationships', or had got the very 'idea' of what was required, may have been generated by alternative unintelligent response strategies. Realising that the speed of learning and the number of trials to criterion were only weak indicators of the possession of cognitive competences, the rationale of Yerkes was to arrange different species along non-trivial dimensions of intellectual ability. He further claimed to have used the existence of "ideational behaviour" to support the existence of qualitative species differences. However, such differences would require cognitive competences beyond those (of albeit complex) rule-stacks which might have explained correct choice in fixed-choice decision-making experiments, and no other contributions which might have helped address the issue of the evolution of intelligence in animals are provided in Yerkes' writings. Nonetheless, as with Köhler, the problem-solving environments set up by Yerkes did allow the experimental animals to control some aspects of their behaviour on the basis of complex spatial and ordinal relationships perceivable between the items of a set of multiple alternatives. This latter point is of important to the discussion of a second continuity versus non-continuity factor for our understanding of the natural evolution of intelligence. In the 'insight' experiments of both Yerkes and Köhler, one can see the need for an animal's recognising the nature of the task environment and its potential component parts across sessions (continuity), whilst simultaneously attending to the specific environmental components and their ongoing spatio-temporal displacements within the course of any given session (discontinuity). It may be seen, therefore, that as well as viewing continuity/non-continuity issues at the level of individual species' competences, one might also apply the same distinctions at the level of learning mechanism. I will return to this point again when discussing

more recent developments (especially the work of McGonigle) in which it is a requirement of the subjects (of whatever species) that they operate with a given set of items in the face of alternative solutions, always having something to do with regards some level of autoregulatory control despite upwards shifting of the task demands. Whether for the earlier or more recent research, those tasks which necessarily require the recombination of known items in novel situations (rather than relatively tightly coupled stimulus-response success) would, should they exist, give rise to better evidence for individual species differences.

Paradigmatic restrictions

Despite their claims to support the case for continuity between the mental processes of animals and man, the earlier body of work reviewed above is found to be clearly lacking in any standard paradigms or experimental procedures. Their observational methods and measures were never refined and the subsequent data sets acquired from such studies remained either incomplete or of an almost entirely anecdotal nature, making their replication almost impossible. In the fifty or so years following Darwin's writing of *The Descent of Man*, the view that the mental life of animals had evolved in much the same way as had their physiological characteristics, was beginning to gain much criticism. The search for continuity could not be satisfied by means of introspective methods alone, and, until quite recently, very few comparative psychologists continued to concern themselves with attempting to infer the mental lives of non-human animals directly, content to merely report what they could and could not *do* (see also later this chapter). As a result of this general loss of direction and interest in mentalism, a new behaviourism was to emerge in the early part of the twentieth century. The motivations for its development and subsequent progress will now be reviewed in so far as it may help to resolve this apparent impasse in our attempts to explain the evolution of intelligence in complex organisms.

The rise of behaviourism

Our discussion has so far presented the progression from Darwin through Romanes to Yerkes as providing a continuous, yet evolving tradition for the study of animal behaviour. At the same time, and in parallel with this growing tradition, another continuous, and evolving line of enquiry was

branching from Romanes through Morgan and Thorndyke to Watson. This second school of thought was primarily motivated by the need to understand the rules or laws by which mentation might be inferable from behavioural observations, an issue still left wanting since the time of Darwin. In his book *An Introduction to Comparative Psychology*, Lloyd Morgan (1894) gave little weight to the kinds of evidence provided by Romanes, believing that even the skilled, detached observer of animal behaviour would contaminate their data with their own preconceptions. It was also in this volume that one sees the first appearance of the oft quoted 'Morgan's canon' (still used today both as axiom of attack against anthropomorphism and as a defence for empiricism):

> 'In no case may we interpret action as the outcome of the exercise of higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale.'
>
> Morgan (1894: p. 53)

The use of persistent and repeated observations over time were also to be prescribed by Morgan, but he did not extend his position to promote a fully experimental approach to the subject. Morgan was really more concerned with the process(es) of learning, and his own observational work with animals convinced him that all reliable evidence for intelligent behaviour could be explained in terms of what he called 'trial-and-error learning'. In his later work, *Habit and Instinct*, applying the Spencer-Bain principle Morgan (1896) would go on to distinguished two kinds of processes, both associative in nature, and, following his studies of newly-hatched chicks, then gave up any remaining belief in the Lamarkian principles of heredity. Once he had decided that no trace of the habits and associations acquired by an individual animal could be passed on in any biological way to its offspring, he then formulated a much clearer idea of what was to be meant by 'instinctive' behaviour. Another contribution of Morgan, was to highlight a further point of oversight on the part of the early continuity school supported by Romanes, a point which was later to become of critical import to our developing story of the evolution of animal intelligence. On the issue of reasoning, Morgan agreed with Spencer's view that reasoning might have evolved from the processes of perception, but as well as doubting the very existence of reasoning in non-human animals, he thought that the origins of human reasoning were closely related to the

development of language. This latter comment was to be predictive of a later division in the study of animal behaviour, and I will return to this issue later in the chapter.

A second criticism of Romanes work was to come from Thorndyke, who disliked the anecdotal nature of the behavioural evidence put forward to support a continuity of mentation between animals and man. In his review article, *Animal Intelligence: an experimental study of the associative processes in animals*, Thorndyke (1898) claimed that too much of a case was being made for the 'intelligence' of various species of animal at the expense of discussing (albeit equally anecdotally) their more frequent 'stupidity'. His own work was grounded in evolutionary theory, reflecting the general background provided by Darwin, Spencer, Romanes and Morgan. However, this was subsequently augmented by the development of a more sceptical attitude and Thorndyke refused to accept anything other than quantitative experimental evidence as contributing towards the resolution of the question of animal intelligence. Indeed, even for a better understanding of habits and human learning, Thorndyke believed introspective reports to be irrelevant or even to hinder such research. His subsequent recommendation was for the design and implementation of more objective methods of observation and experimentation. Thorndyke's more rigorous experimental procedures resulted in his formulation of stimulus-response theory, an explanation intended to provide a simple solution to what was then known as the 'performance problem'. This idea was later to be stated quite explicitly in his *law of effect,* in which:

> 'Of several responses made to the same situation, those which are accompanied or closely followed by a state of satisfaction to the animal will, all other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur.'
>
> Thorndyke (1911: p. 244)

By this reasoning, even if an animal could be said to have learned an association between a response and its outcome, one still required an explanation as to how such learning could be translated into performance. No such problem existed if, as Thorndyke's believed, the only function of a reward was to 'stamp-in' a connection between stimulus and response, since the occurrence of the appropriate stimulus or situation directly 'called up' the connected response. He thereafter came to believe that the

essence of intelligent behaviour lay in the formation of associations, and that species differences in intellectual capacity reflected quantitative differences in the number, and speed of formation of such associations that they could demonstrably make. With Thorndyke, the purpose, as well as the methods for studying animal psychology began to change. Furthermore, by the time he had failed to find much evidence of imitation or passive learning among the cats, dogs and monkeys that he tested (see Thorndyke, 1911) he had also begun to suggest the existence of an even wider separation between human and animal intelligence than had been proposed by Morgan, (also having found the difference *between* the non-human animals to be relatively small, as indexed by learning curves obtained from his instrumental conditioning experiments). But although Thorndyke failed to obtain any evidence of the capacity to reason in monkeys, he nonetheless argued that:

> '... rational connectives are, in their basic causation, like any others, the difference being in what is connected', so that, 'the denial of reasoning need not mean, does not to my mind, any denied continuity between animal and human mentality.'
>
> Thorndyke (1911: p. 294)

A new methodology

Having now reflected upon this first decade of the current century, it is now apparent that a new line of enquiry was being argued for in the study of animal psychology. Receiving support from biologists (e.g., Loeb, 1911) and comparative zoologists (e.g., Jennings, 1906) alike, it was becoming widely agreed that one should no longer attempt to draw any inferences about an organism's subjective experience from its behaviour. Animal behaviour was now starting to be framed in terms of materialistic explanations, and the notion of consciousness as causal agent (although still prevalent in the humanistic psychology of the time) became excluded entirely from the analysis of non-human animal behaviour. It was becoming clear that simply the species-comparative investigations of sensory capacities, extent of trial-and-error learning, degrees of imitative abilities, and so forth, were not leading to any progress concerning animal mentations. Watson (1913) was perhaps going too far in claiming that all of the 'behaviour men' had come to recognise this point, but the majority were by now well along the way to doing so.

By 1910, the two lines of enquiry had already begun to diverge. Watson moved in the direction of a new behaviourism, whilst Yerkes remained explicitly situated in a comparative psychology born of the evolutionary tradition. Despite attempting to conduct more systematic experiments which might yield more quantifiable results, Yerkes nonetheless continued to interpret them within the framework of Romanes. And although the work of both Yerkes and Köhler had become quite widely known prior to the work of Pavlov (not widely published in English until 1927), their influence remained limited.

In an experimental behaviour-based physiology laboratory, Pavlov had been using a fundamentally different procedure in order to study learning in animals. In the absence of any problem-solving task as such, a typical experiment might involve the restraint of a hungry dog in an experimental chamber (see Figure 1.4) and its salivatory response to the presence of food, or associated events being measured.

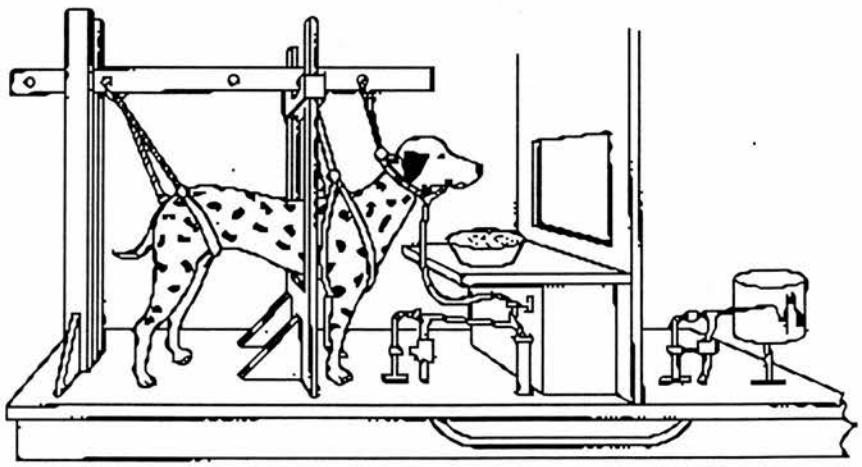


Figure 1.4 Apparatus used by Pavlov for his classical conditioning studies. (adapted from Yerkes and Margulis, 1909)

Essentially anti-mentalist, Pavlov's (1927) work with the mammalian digestive system resulted in his description of the phenomenon of Pavlovian, or classical conditioning: a change in an animal's behaviour resulting from a strictly temporal relationship between two events. Such behavioural changes were said to be most marked when the first event was of little initial interest to the animal, but shortly preceding another event that was of considerable significance. This phenomenon had long been

known in a general kind of way since Spencer (1855) had given conditioning a prominent place in his earlier thoughts on mental evolution. However, having drawn heavily from Pavlov, Watson (1913) was to shift towards the position of Thorndyke, acceding to the view that animal intelligence remained limited to the acquisition of habits. He refused to accept that any process more complex in nature than 'stimulus-response' connectionism could be entertained without positing a dualist way of thinking. Indeed, as Boakes (1984) has pointed out, to discuss 'ideation' in racoons as Cole (1907) and Hunter (1914) had done, was to allow the soul back into psychology. This would be anathema to the later Watson, who came to deny even to the human, any thoughts or feelings that could not adequately be analysed in terms of some determinable peripheral stimulation (see, e.g., Watson, 1913, 1924). Although not explicitly stated by him in his writings, it would appear that this singular idea was to underpin the rationale for the new behaviourism of Watson (and his subsequent followers), promoting a strict methodology for an experimental science of behaviour with a view to providing more objective, quantitative, reproducible data for analysis. It is unclear how quickly these behaviourist ideas spread through psychology, but by the mid-1920s, Watson was recognised (at least in the USA) as being one of the leading American psychologists of the day. From this time until the early 1930s, behaviourism was in an attractive position because it seemed to assure that psychology could at last be grounded in a more scientific methodology. Whilst at first glance appearing to be progress, and although most psychologists at the time appearing to accept Watson's methods, they did not share his vision that psychology be relevant to real life. As a result, and as he became less research active himself, so behaviourism became narrower in focus than it might have remained under his continuing influence.

Operant conditioning

Extending, and later to become a leading exponent of, behaviourism, B. F. Skinner arrived at Harvard in 1929 to work for his doctorate, committed to this new scientific and practical psychology. However, his first major theoretical contribution had tended to go against Watson's ideas, following his dissatisfaction with basing everything on the reflex:

> 'The attempt to force behaviour into the simple stimulus-response formula has delayed the adequate treatment of that large part of behaviour which cannot be shown to be under the control of eliciting behaviour.'
>
> (Skinner, 1938: p. 20)

Indeed, it would appear to him that not only did people respond to the environment, but also that behaviour 'operated' on the environment to generate consequences. In contrast to the instrumental learning of Thorndyke discussed above (in which the hungry cat in the puzzle-box would learn to escape, and thereby gain access to food), Skinner was to devise a similar technique to test his 'operant conditioning' idea. This new technique typically employed a hungry rat or pigeon in a "Skinner box" (at the time a very modern technical apparatus) who were not this time required to determine a method of escape, but were merely required to press a lever (electrically coupled to a feed dispenser mechanism) in order to receive food delivered to them whilst remaining in the box.

In this new situation, if the rat received food every time it pressed the lever, it was said to have 'operated' on the environment. Whereas for Thorndyke the animal was no longer available for comment once escaped, for Skinner's subjects, the post-fed animal continued to be ready in a position to respond once more, so providing a convenient measure of learning in terms of changes in the rate at which lever-pressing might occur over the course of the ongoing session. Skinner referred to this procedure as 'operant conditioning' (the animal operating upon its environment in order to have the effect of obtaining its food). Rather than viewing the food rewards as reinforcing stimulus-response associations (as for the *law of effect* discussed above), Skinner's approach was not to theorise about it, but to use it. For him a 'reinforcer' was any event that brought about a change in the likelihood of the behaviour upon which it was consequent. There was little discussion concerning the changes in the strength of stimulus-response associations in Skinner, but instead, reinforcement was presented as a technique for the control of behaviour. Reinforcers may have taken the form of food presentations to the hunger-motivated animal, but, as Skinner was to find, so too was the presence of a light that had previously been paired with food. In this latter case, when the lever-pressing maintained the illumination of the light, the rate at which pressing would occur also increased. Skinner

(1939) was then to report on a whole set of investigations designed to disentangle different aspects of learning using this new technique. For example, in determining the effects of extinguishing a reinforcer, he distinguished two kinds of positive effects: those allowing an animal to escape an unpleasant stimulus (negative reinforcement), and those allowing access to a pleasant reward (positive reinforcement). The important feature here was that for the former case, the learned response was particularly resistant to extinction, because the animal never gave itself an opportunity to learn that the unpleasant consequence was no longer going to happen.

In this and later work, Skinner (1938, 1969) found that it was possible to alter the resistance to extinction of learned behaviour by manipulating the reinforcement contingencies (i.e., adjusting what the animal was required to do in order to gain a reward). In an attempt to distinguish between the discrimination of stimuli and a process of differentiating between forms of response, he found that this type of behaviour could become quite resistant to extinction if trained in the appropriate ways. The principle methods of testing for contingency detection involved the administration of different schedules of reinforcement (continuous, and fixed Vs varied intervals and ratios), with each found to have a distinctive associated outcome. This process of 'behaviour shaping' is important here (and to the experimental component of the current thesis) for three reasons; firstly, it put forward a learning mechanism whereby entirely novel forms of behaviour could be constructed in the laboratory (classical conditioning could only deal with an animal's existing responses). Secondly, the behaviour shaping dealt with the training of voluntary behaviour, not merely the autonomic functions. Thirdly, large quantities of data could be reliably accrued over long periods of time, with the same animals being monitored during their performance with different tasks along their individual developmental trajectories, the life-history of the subject being continually profiled. What Skinner argued for, was the belief that his behaviour shaping was the principal process by which both humans and animals learned more complex kinds of behaviour. Indeed, in a more extreme statement, Skinner (1957) was to suggest that even human language could have developed as a result of parents (and others) reinforcing the child's operant behaviour of babbling, such that it became shaped to include an increasing number of word-like sounds,

until eventually the child was able to produce whole words, and then sentences. This issue will be returned to later in this and the subsequent chapter, but for now, it is offered as an example of Skinner's more extreme behaviourist outlook, in which he appeared to see *all* human behaviour as potentially explicable in terms of ongoing conditioning processes.

Interim summary: evolutionary to behaviourist stances [1850 - 1945]

Our discussion of the rise of behaviourism to the forefront of comparative animal psychology in the early half of this century, has seen its success to have been motivated by the need to produce a more quantitative, empirical science of behaviour. Despite Spencer's preempting what was to become classical conditioning, it was not explicitly investigated prior to the detailed work of Pavlov (1927), and in hindsight, the loss of interest in Yerkes & Köhler's work appeared to be due largely to Pavlov's more familiar kind of theory and objective methodology. But more importantly perhaps, taken forward by Thorndyke and Watson, it was easy to see what kind of empirical questions needed to, and indeed could, be answered within the framework offered by the conditioned reflex, instrumental, classical, and later operant conditioning theories. In contrast, it was not at all clear what kind of research might have developed naturally and productively out of the earlier studies of problem-solving. It is unclear how quickly these behaviourist ideas spread through American psychology, but by the mid-1920s, Watson was recognised as one of the leading American psychologists. The subsequent years of study within this new and more rigorous experimental field of behaviour was also to set seal upon the issue of the continuity of mentation between man and the other animals. Following the lack of evidence being provided for the existence of clear quantitative differences in the performance of different species in purportedly identical laboratory learning experiments, and in a climate of rejection of consideration of the value of any apparent qualitative differences, a shift towards a discontinuity position was beginning to be felt. A student of this more 'scientific' behaviourism, Skinner remained essentially anti-mentalist, ascribing to the view that it was an animal's history of reinforcements that determined its behaviour. In his final discussions, Skinner (1938) admits having little to say to the neurologists searching for the physiological subcomponents of higher mental processes such as cognition or consciousness, and remained content to

view them as being merely epiphenomena. For Skinner, feelings were not the causes, but the consequences of behaviour, and behaviour could be predicted and controlled without reference to them.

Although unable to support the view of there being any evolutionary continuity between the cognitive processes in man and other animals, there was, however, another contribution provided by Skinner's new technique that was to have lasting important consequences for the development of studies in comparative animal learning. In reusing the same individual animal subject in his subsequent shaping and reshaping experiments, Skinner was necessarily taking into account the life historical experience of his subjects as they progressed through their various schedules of reinforcement. This factor was to make its expression in the work of Harlow (1949), who investigated the possibility that trial-and-error learning, based on operant conditioning techniques, might lead to apparently 'insightful' behaviour. Further, and without the need to invoke cognitive explanations, whatever the outcome of individual experiments, he thought that he might also be able to reveal any extant species differences in the process.

Learning to learn: a task-informed and life-historical approach

The importance of Harlow's work for the story here is that, as well as having provided an attempt at avoiding the restriction of simple conditioning experiments, he had developed a new technique, using the Wisconsin General Testing Apparatus (WGTA), a device designed to allow species' comparisons (of higher cognitive behaviour) that did not rely upon simple associative learning for task success to be demonstrated (see Figure 1.6). Although his learning set experiments (see below) were not extended, nor his 'learning to learn' interpretation fully characterised, Harlow's work was to help keep alive a non-associative interpretation of animal learning throughout the following years of fierce behaviourist attack. As discussed previously, the efficient solution to the sort of problem set by Köhler (1925) was dependant upon the opportunity of the animal for the development of specific motor sequence skills, if only in the course of play. For example, one implication was that Köhler's chimpanzees may have been simply learning which responses were instrumental in obtaining bananas, and that the responses in question formed part of their natural repertoire. This lent itself readily to conform to the ideas of

instrumental conditioning and to Thorndyke's interpretation of his 'puzzle box' experiments: responses (or patterns of responses) may have been selectively reinforced by their associated consequences, forming sub-goals on the route to obtaining a banana (see discussion p.7 above).

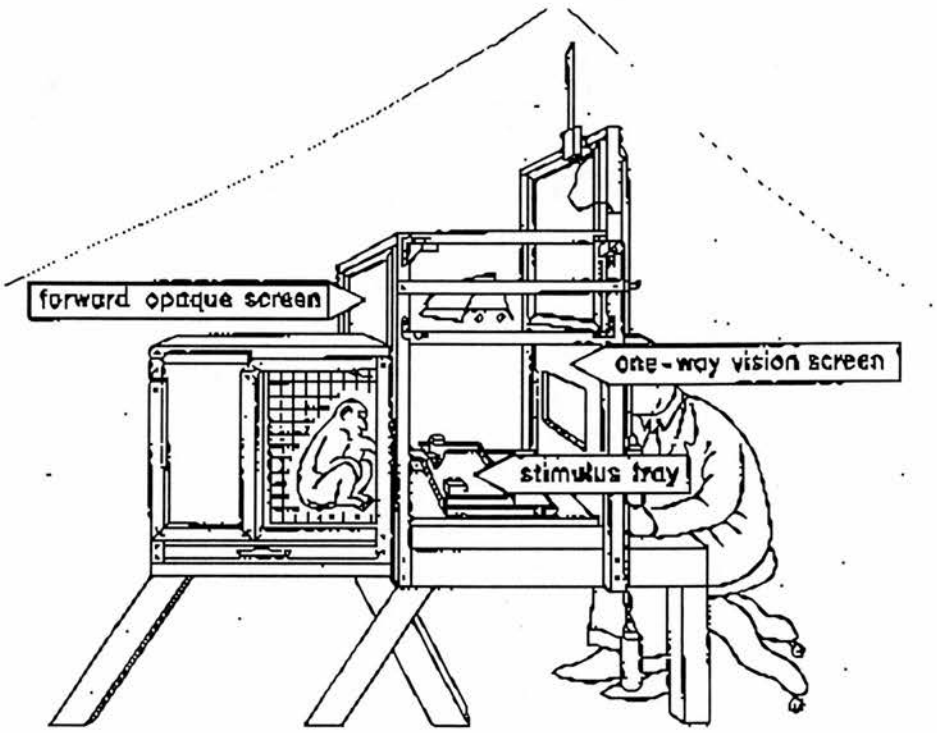


Figure 1.6 Wisconsin General Test Apparatus (WGTA) (Drawing from original. Harlow, 1949)

However, whichever way Köhler's results are viewed, his experiments nonetheless established that an animal's past experience could have profound effects upon its subsequent ability to solve certain kinds of problems. This conclusion was well supported and illustrated in the work of Harlow (1949) on learning sets.

In one experiment, Harlow trained rhesus monkeys on simple two-choice discrimination problems. The animal was confronted with two different objects placed over two small food wells. Displacement of one of the objects revealed, say, an obtainable peanut or a raisin, in the food well, the other well being empty (see fig. 1.7 (a) & (b)).

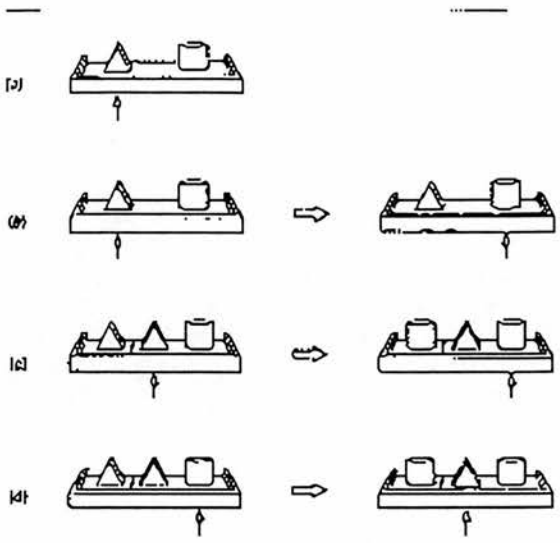


Figure 1.7. Harlow's WGTA presented tasks: (a) visual discrimination (b) visual discrimination reversal (c) Matching (d) 'odd-one-out or non-matching.

The monkey's task was to learn which object, independent of position, concealed the reward. Discrimination could be interpreted as a case of conditioning - one object or stimulus was indeed associated with a reinforcer and elicits approach, the other was not, and did not. But although this analysis may account for the way in which an animal might have solved the first discrimination problem that it was set, Harlow did not stop at that point. Once the monkey had solved one problem, it was set another: two new objects were presented, and the monkey had to learn all over again which one was 'correct', and so on. Whereas the animal had initially taken maybe a dozen trials to learn which of the objects was the correct one, after more than a hundred trials, their behaviour had dramatically changed. As may be seen from fig. 1.8, with growing experience of the different problems, the rate of success for the subsequent trials on each stimulus set showed a progressive improvement. Not knowing which alternative was correct on the first trial of a new problem, the animal would choose between them at random and could be correct fifty percent of the time. But on trial two, they seemed to "know" the solution and always chose correctly. In Harlow's phrase, the animal was '*learning how to learn*', and was said to have formed a '*learning set*'. (Harlow, 1949: p53).

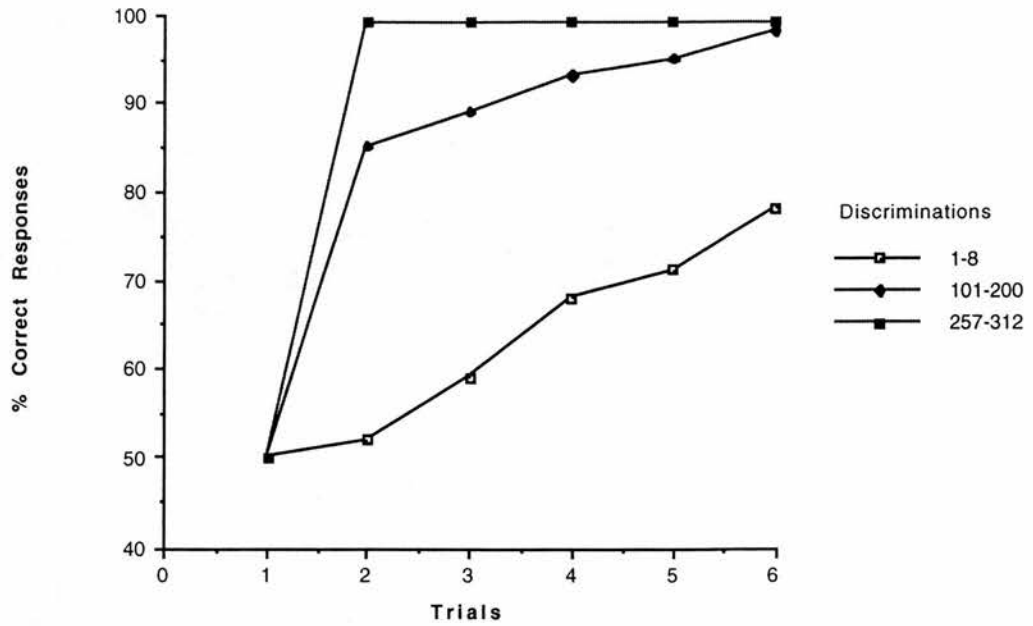


Figure 1.8 Discrimination learning curves on successive blocks of problems. (adapted from Harlow, 1949)

To explain this improvement, it was not possible to put forward the simple associationist account, as this would only be effective when the same pair of stimuli were used throughout the experiment. Nonetheless, a slightly more complex version has been used to interpret Harlow's findings. Accounting for these results by proposing the animal use a 'win-stay, lose-shift' strategy on a trial-to-trial basis, Restle (1958) had suggested that the animal was learning simply to repeat choices following rewarded choice. However, this did not question the associative basis of performance on learning sets. The experienced monkey was still assumed to be associating one alternative with food and the other with its absence, and to choose between them accordingly. Because it is quite difficult, although not impossible (see also Reese, 1964) to interpret learning sets in terms of associations between particular stimulus-response events, one might see for the first time the emergence of claims for higher-order learning or rule-learning in a non-human animal. Together with other experiments involving more complex tasks such as 'odd one out' and 'repeated reversal' problems (see Figure 1.7, (b) & (d)), the learning set results continued to defy stimulus-response explanations, leading to a number of comparative

psychologists concluding that the animals were not just learning the problems faster, but were acquiring the 'principle' or 'concept' behind the problem (e.g., Hays, 1994; Pearce, 1987).

What does the learning set measure?

Whatever interpretation is to be preferred, although Harlow argued for an unifactor continuity position, his learning theory is based on inhibition only and it is still unclear what exactly he believed his animal to have learnt. His connectives remained arbitrary and were presented over a limited search domain, giving rise to small error spaces for interpretation in the absence of increasing task difficulty. Furthermore, although Harlow (1949: p. 56) argued that:

> 'the learning set is the mechanism that changes the problem from an intellectual tribulation into an intellectual triviality and one which leaves the organism free to attack problems of another hierarchy of difficulty'
>
> (Harlow, 1949: p. 56)

... he did not have at his disposal the means to objectively establish a hierarchy of task difficulty. As a consequence, he was unable to offer any evidence or advise on the means by which the transfer value of a learning set could be evaluated in a different - and more difficult - problem domain.

What Harlow's procedure did allow, however, was the observation of a monkey demonstrating the benefits of multiple problem solving as assessed by a brand new measure of inter-problem transfer and not just measures based on single discrimination learning problems per *se*. Reasons were thus being provided for a distinction coming to be made between simple conditioning and more complex forms of learning. The complexity for the learning set resides in the kinds of stimuli to which the animal is required to respond, and to the ways in which they might analyse or categorise the stimulus array, rather than to simply respond to the putative degree of complexity in their relationship to subsequent events of consequence to the animal. If indeed such 'higher-order' learning is involved, as McGonigle and Chalmers (1997a) might be taken to imply, a task testing the ability of an animal to make more abstract

generalisations, or rules, might lead one to expect that any species differences found might conform to our more intuitive estimates of their intelligence. However, extensive use has more recently been made of Harlow's procedures with a variety of species, and comparative studies of learning set formation have been conducted. Passingham (1981), reviewing this work, has pointed out that primates learn these rules quicker than do non-primate mammals such as rats, cats or squirrel (see fig. 1.9). Furthermore, rhesus were found to far outperform the small New World monkeys such as marmosets, and rhesus themselves were outclassed by the chimpanzee. Passingham reviews other rule-learning tasks, showing that chimp, rhesus, squirrel monkey, marmoset and ring-tailed lemur could all learn the rules, but that their facility to do so decreased in that order (Passingham, 1982).

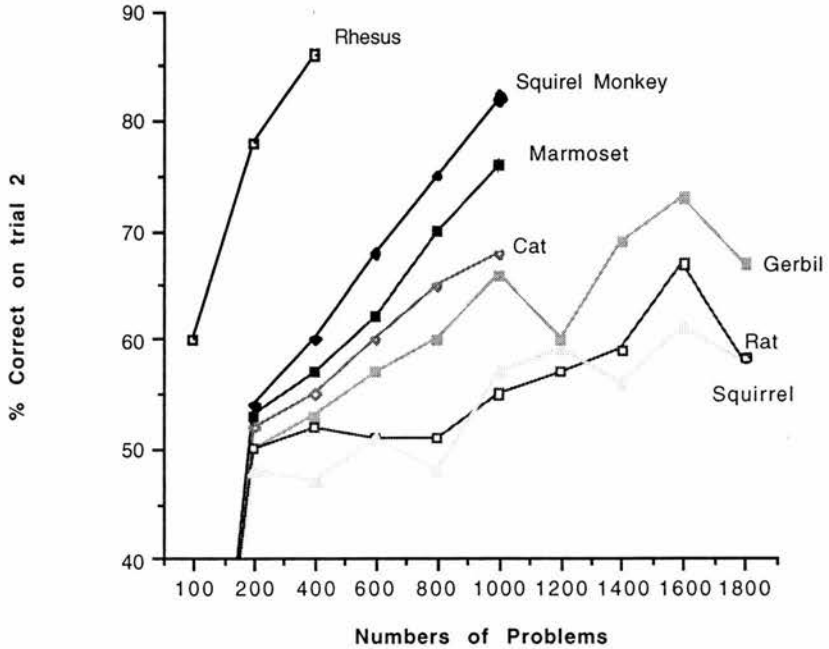


Figure 1.9. Learning set formation in six mammalian species. (After Passingham, 1981)

The "levelling" hypothesis: the case for discontinuity

After reviewing the vast corpora of data derived from the studies of instrumental and Pavlovian conditioning over the last 100 years, Macphail (1982) concluded that all vertebrates (with the exception of man) were of equal intelligence: his Null Hypothesis. Furthermore, Warren (1965),

having reviewed fifteen years of research on learning sets, was to report that:

> 'Mammals and birds differ from fish and reptiles in being able to learn repeated discrimination reversals in progressively fewer trials, but primates are not markedly more proficient than other mammals and birds.'
>
> (Warren, 1965: p. 266)

Although not denying the possible convergences of species-specific responses to niche-adapted complex environments (as typically described by the mid-20th century ethologists, e.g., Lorenz, 1965 and Tinbergen, 1951), Macphail nonetheless preferred the study of animal behaviour to be confined to situations for which the demands made upon the subject allowed it to "demonstrate the behavioural flexibility that is implied by such intelligence as it might possess". Once again, one sees the emergence of a most noble aim, but what situations might be so designed ? What might be the *a priori* conditions for a 'good' experiment ? How were the 'levels of difficulty' to be evaluated for a given species ? On these issues Macphail has little to contribute, but for complex learning situations (non-habituation, classical or instrumental conditioning) he makes the following comment:

> 'Where a particular phenomenon can be demonstrated in one group, but not in another, this will be taken as *prima facie* evidence for a qualitative difference between those groups; it will not, of course, be taken as good evidence for such a difference unless accounts in terms of plausibly relevant contextual variables are ruled out through systematic variation'.
>
> Macphail (1982: p.14)

His writing does suggest, however, that in principle, if there were to be further evidence for comparable quantitative differences between the same species in another task posing quite different demands, then support for an 'intellectual' interpretation would thereby be strengthened.

In contrast to this view, Maier and Schneirla (1935), writing in their volume *Principles of Animal Psychology*, concluded from their own experimental work that the abilities to form associations and to demonstrate reasoning could be made quite distinct, and that to operate the latter, a well developed cerebral cortex was required (and to be found only in mammals). Other researchers had previously looked to brain activity

correlates in search of an explanation for differences in behaviour with mixed success. The two most prominent allometric variations across vertebrate brains were those reported to be of (i) overall size differences and (ii) the relative size and differentiation of specific brain regions (see e.g., Byrne, 1993, Jerison, 1969, 1973; Russell, 1979; Dunbar,1997). Lashley's (1929) lesion studies with rats working on a complex task (maze-learning), had concluded that the neocortex (at least for the rat) be equipotential, although the more intact the cortex, the more efficient the performance remained. For all of the comparative allometric data, whether one studied normal, experimentally induced, or differentially brain-lesioned individuals, it remained open to question as to what anatomical and functional differences might underlay any of the observable behavioural differences previously detectable. (A hope does still remain for the future, however, with the recent development of less invasive brain-scanning techniques).

One's optimism concerning the utility of any of these approaches will necessarily relate to one's position on the continuity issue discussed so far above: if there are to be many diverse mechanisms involved in intelligence, then it is reasonable to suppose that not all of these mechanisms are going to be common to all species. Similarly, at least some of these mechanisms might be related in some orderly way to some specific brain regions. It will therefore, come as no surprise to note that a physiological psychologist such as Lashley (1929) and comparative psychologists such as Maier and Schnierla (1935) had argued that the principles of associative learning failed to provide adequate explanations for all forms of intelligent behaviour. Conversely, learning theorists in the Thorndykian tradition had assumed that there were no important species differences, and thus viewed physiological experimentation with reservation. When discussing species differences, Hull was to write that:

> 'the natural-science theory of behaviour being developed by the present author and his associates, assumes that all behaviour of the individuals of a given species, and that of all species of mammals (including man) occurs according to the same set of primary laws.'
>
> Hull (1945: p. 56)

Similarly, Skinner was to claim that:

> 'I may say that the only differences I expect to see revealed between the behaviour of rat and man (aside from the enormous differences of complexity) lie in the field of verbal behaviour.'
>
> Skinner (1938: p. 442)

This comment from Skinner was a restatement of that put forward by Morgan (see p.10 above) towards the end of the previous century, and one reads it once more in the writing of Macphail concerning the differences between human and non-human intellectual capacities.

Complex learning abilities and language

Macphail (1982) was also to hold the view that the 'superiority' of humans (whether qualitatively or quantitatively measured) might simply reflect the possession of language. This single factor had by now provided the basis for much discussion and the view that species-differences in learning mechanisms could be explained by the absence of language in all mammals except man, was to become more widely accepted. For example, having pointed out that animal intelligence should not be judged by comparison with human intelligence, this was said to be because human intelligence had become special, as a result of:

> '.... language and related skills, which permit us to communicate not only with each other, but with past and future generations.'
>
> Hodos (1982: p. 37)

If, as one might consider, the levelling hypothesis was to be upheld, and if the behaviour under review was not to be extended in any way, Macphail (1982) was to correctly pointed out three candidate scenarios presented themselves for further comparative investigation:

(i) humans possessed the same complement of learning mechanisms as did other animals, but as a result of their quantitative superiority, allowed language acquisition. [humans do so because they are more intelligent]

(ii) humans possess some complement of learning mechanisms as did non-human animals, *not* quantitatively different, but *also* possessed an additional mechanism(s) who's sole function was language acquisition. [humans are more intelligent *because* they possess language]

(iii) humans possessed some compliment of learning mechanisms not qualitatively different from those of other animals, but in addition

possessed a mechanism(s) which allowed language acquisition *plus* the solution (in some novel way) to other problems. [humans are, independent of language, more intelligent than other species]

Although helping to give orientation to the comparative psychologists emerging from fifty years of a strict behaviourism resulting in an equivalence position on the issue of vertebrate mental abilities, these questions were not of themselves to provoke any new procedures or paradigms for their being distinguished from one another. Indeed, even Macphail himself was only to continue the discussion with a review of the language acquisition work with apes (e.g., the works of Kellog, 1933; Gardener & Gardener, 1969; Premack, 1971; Rumbaugh, 1977; Terrace et. al., 1979) after which he concluded that no convincing demonstrations of multi-sign strings had been produced that resembled the sentences of human language in any critical sense (see also next section below). Despite his inability to state that non-human species could *not* acquire a language, Macphail (1982) proposed that man had acquired language (and that other species had not) not because humans were quantitatively more intelligent, but because man was possessed of some species-specific mechanism(s) which provided a necessary prerequisite for the acquisition of language. The best known proponent of this view was Chomsky (e.g., 1972), who had reached the same conclusions but for very different reasons (see also next chapter). Despite the apparent impasse to have been provided by Macphail's (1982) 'null hypothesis' for the resolution of continuity factors between species in intelligence, he also argued that when a difference *was* revealed between two species on a task, that this might then be the result of poor experimental design rather than to the existence of any real differences in intelligence. This, however, is logically difficult to disprove. Furthermore, for Macphail, any differences that were to be found (be they qualitatively or quantitatively measured) were said to only reflect the possession of language. It is to this topic that we now turn, providing a review of the work that has been conducted with non-human species (especially the great apes) in attempts to determine the degree to which, if any, these subjects might reveal themselves to be at all capable of utilising a human-like language facility.

Non-human animal language, logic and thought.

Communicatory signals as exhibited in the wild by a wide variety of animal taxa are extremely common, but it would appear that species other than humans are not naturally disposed to use our kind of language system. Furthermore, without the observation of some equivalently vocal or richly symbolic, syntactically structured nuances as can be seen in human verbal communication, it has been very difficult to determine whether some species might be otherwise capable of demonstrating rational thought. Some researchers (e.g., Hayes & Hayes, 1951; Kellogg, 1968) have explicitly attempted to teach animals our kind of language facility, others (e.g., Herrnstein, 1979; Schusterman & Gisiner, 1984; Premack, 1976) have continued to use traditional methods to determine whether animals might be capable of demonstrating instances of categorisation and concept formation. The successes of these research programs, as well as their better known failures, have continued to provide much experimental data for consideration with regards the possession of cognitive adaptations by non-human species.

Common to all of the paradigms so far discussed, there has been a continued need for some form of behaviourism or behaviour-based analysis which has remained essential for the evaluation of cognition. This is true for human, non-human and non-biological systems alike, and what animals actually *do* will for the present have to be the final test for what they *think*. Furthermore, the degree to which a given animal can be said to perform abstractions *of any kind* are, next to language, the most popular dividing line between animals and man (Craik, 1943). However, again, the most salient indicator of the presence of abstraction in the human, has been its explicit association with verbal labels and their use in language and thought. It would appear therefore that another impasse has presented itself along our quest for an answer to the continuity question concerning the mental life of human and non-human animals. One further avenue of approach might, however, be to determine whether there is any evidence for some instantiations of *pre-linguistic* thought in the animal kingdom.

In order to test the hypothesis that there might be such a thing as pre-linguistic thought, one might first attempt to examine the psychological

processes of young children (see McGonigle & Chalmers (1980) and later chapters below) in whom ideas might occasionally 'run ahead' of the developing adult-like speech. This could be assessed through the investigation of perception and action, if such were to provide evidence of classification, categorisation and the compounding of ideas, or the activation of inner memories and expectancies. Procedures so designed might lead to the manifestation of non-linguistic mental events in a way analogous to adult human speech. However, before embarking upon our review of the claims for gestural and symbolic communication skills in the great apes, it is pertinent here to mention the results of a few non-primate investigations concerning categorisation and logical thinking in other species.

Working with the pigeon, Herrnstein (1979) has reported natural category learning using the presentation of coloured transparencies as stimuli for discrimination. In one experiment containing eighty slides, forty of which contained trees, only pecking to a slide depicting a scene with a tree was reinforced with food reward. After a relatively large number of trials, most pigeons were able to reliably discriminate tree versus non-tree slides. This ability to form a category did not remain limited to the identification of trees (an object of obviously adaptive significance to the bird) but was later shown to extend to slides of people, fish, bodies of water and even alpha-numeric characters (Herrnstein & de Villiers, 1980; Morgan et. al., 1976). Furthermore, not only do these studies provide us with reasonable evidence for the existence of a conceptual grouping ability in the pigeon, they also suggests that although categorisation is required for human language, language of the type possessed of the human is not a necessary precursor to categorisation. In a similar vein, it might be conceded that logical thought not require language either, and that language is only needed for *discussions* of logic and logical argument. Indeed, only since the writings of Piaget have psychologists thought to deny that language develops prior to cognition (see Piaget, 1971, and Chapter 2 below).

Using simple logic problems which simultaneously require category formation, Schusterman & Geisiner (1988) have shown sea lions to be capable of demonstrating syllogistic thought. In their experiments, these marine mammal subjects were trained to first choose a symbol B rather

than an alternative when A had been presented immediately beforehand. They were then rewarded for choosing A following presentation of B; and reinforced for selecting C following presentation of B, and vice-versa. Finally, in the test phase comprising untrained choice combinations, after presentation of A, the sea lions immediately selected C rather than an alternative. Another task involving training on X=Y and X=Z, testing for X=Z ?, also showed this species to be competent at such logical tasks. This novel technique developed by Schusterman is yet to be tried with any other species (and other techniques have failed to show such performance with any animal), but it does inform us that perhaps for any given animal, it is perhaps one's approach to training which builds upon the animal's natural frame of reference that might be the key to reaching the animal mind. However, evidence for symbol-based syllogistic thought is a long way from the spontaneous production of sequencial structures more typical of human language.

Although one is not yet in a position to interview non-human subjects with a view to establishing the nature of any mental operations which they might undergo, attempts have nonetheless been made to teach human vocal sounds to birds with a reasonable degree of success. In particular, members of the parrot family will often reproduce good approximations to a large number of sounds in their environment, including human speech, even in the absence of explicit training. The question to be asked, however, is whether these 'talking' birds might be capable of demonstrating any understanding of anything concerning the sounds which they make ? With this question in mind, a long-term study was embarked upon by Pepperberg (1981, 1983) in an attempt to produce evidence for the degree of comprehension (if any) that a talking African grey parrot (Alex) might have of its utterances. After some ten years of continuous training and testing, Alex proved to be in possession of a seventy-word English vocabulary that included about thirty object names, seven colour adjectives, five shapes, several material names, five numbers and the words for 'colour', 'shape', 'matter', 'what', 'same', different', 'none', and 'no'. Following the acquisition of this very rich word-association sound archive, Alex was able to correctly identify a variety of objects and/or their shape, colour and material properties when asked (verbally) to do so with 95% accuracy, whilst replying using his learned English vocabulary. Even more convincing are his successes with novel,

previously unseen objects, with which no drop in performance levels have so far been recorded (using the same criteria and vocabulary). This latter finding would seem to suggest that the parrot does indeed understand something about the concepts and the logical connections involved.

Taken together, these non-primate examples of experiments concerning the ability of animals to form conceptual categories and to use logic, show that their presence may be determined in the absence of language, and, that they are probably indicative of those species natural adaptations to the most challenging problems they face in their habitats in the wild.

Gestural and signing-based language in great apes.

In both their natural and captive environments, there is no convincing evidence to date that any of the non-human apes show an interest in imitating human speech patterns. Given the nature of the environments within which they interact and evolve their adaptations, it should not perhaps come as any surprise that this be so. Of our close cousins in the ape family, the orang-utan are largely solitary - making little use of any frequent vocal signals; the gorillas are also fairly quiet - using chest-beating as much as vocally produced sound (Schaller, 1963). Even for the Chimpanzees, although more vocal and actively prosocial, it would appear that the subtleties of gesture, posture and touch are on the whole more widely used than vocalisation. Indeed, of all the primate family it is in both the Old and New World monkeys that social vocalisations would appear to be the more extensive (Seyfarth et. al., 1980; Tenaza & Marler, 1977).

It was, however, the apes that were initially chosen for investigation, and by the 1970s, both the Hayes (1951) and the Kelloggs (1968) had reported on the gradual development of problem-solving and the discriminatory abilities of their young chimpanzees reared in their own homes with the hope of observing the appearance of human-like speech. As only a few imitative vocalisations were to be forthcoming (the best was Vicki, who reliably uttered the four single words 'mama', 'papa', 'cup' and 'up'), their efforts have been regarded as largely wasted, although their remains some interest in the non-vocal abilities of the animals as recorded and preserved in various film archives. This latter point is important to bear

in mind, and it was to be the case that the ensuing series of experiments to be conducted with this species concerned with the development of gestural communication, was also to lead to other possible indicants of intelligence being ignored (e.g., Köhler's (1925) observations already stood as evidence for sophisticated spatial memory and object manipulation in the chimpanzee).

Following the failure of these early attempts to elicit the emergence of speech in the chimpanzee (and presumably having observed the successful use of hand-signals and gesture in the training of performing animals) a novel procedure was developed for use in the same species by Gardener & Gardener (1969). Using a gestural language based on standard American Sign Language (ASL), again taught in a social context, a chimpanzee called Washoe was claimed to have acquired a working vocabulary of some 130+ words after four years. The Gardener's analysis of both their real-time and film archive data led them to believe that their apes were clearly able to demonstrate a true understanding of the (often non-arbitrary) connections between abstract gestures and the objects, properties, individuals, attributes, and the action or events that they came to signify. Similar claims were later to be made for a few other species' individuals, especially in the context of simultaneous verbal speech on the part of the trainer during acquisition and interaction, including an orang utan (Chantek) who was said to have mastered some 50 word-signs (Miles, 1983, 1990) and a gorilla (Koko) producing more than 200 (Patterson, 1978).

However, little evidence was forthcoming with regards the production of any grammatical organisation amongst the apes' gestural output and sequences of strings were rarely extended beyond two or three clearly identifiable signs without interruption or repetition. Furthermore, close examination of the photographic records made available of the signing animals (including Washoe and Nim) suggested that they were typically imitating the recent signs made by their trainer who was consistently providing gestural cues during their interaction, albeit unwittingly (Terrace et. al., 1979). There remains, however, sufficient evidence from some of the more controlled studies that there can be little doubt that individual gestures can, and are, reliably associated with particular contexts and events. The disagreements are not so much concerned with whether or not the chimps learn to associate particular gestures with

particular objects or actions, but concern whether the animals are achieving the higher levels of mental organisation that would be indicated by the claim that they are understanding the relationships between the 'words' being used. It is quite possible that this unresolved issue will remain for as long as the technique is unchanged and, as suggested above, perhaps it is the training paradigms and procedures employed throughout these studies which fail to equip the animals in the use of our mixed inflected/word-order system, leaving them bemused as to the intended object of the ultimate tasks being later presented ?

Symbol-based language in great apes.

In the 1970's a quite novel approach was developed by Premack (1970, 1971, 1976) in a further attempt to '..... operationally analyse language, to decompose it into atomic constituents and to provide [chimpanzee] training procedures for each constituent' (Premack, 1976, p.1). His new training system was, he claimed, capable of allowing his subjects to both comprehend and express propositions such as 'these two objects are the same/different' and later 'Sarah give apple Mary'. This interpretation, however, was entirely dependant upon the translation of fairly simple actions upon object arrays into verbal terms. The equipment employed involved the manipulation of small metal-backed plastic shaped objects (tokens) of different colours and sizes, and their placement onto a magnetic board in various ways. Working initially with four chimpanzees, only one (Sarah) proved to be adequately and consistently engaged with the experimental situation to allow the studies to progress very far.

In the studies carried out by Premack (1971) with Sarah, each token is said to 'stand for' a specific object (*apple, pail*), for actions (*give, take, insert*), qualifiers (*red, yellow*), concepts and conditionals (*same, different, colour-of, if-then*). Systematically trained to pick up and place the tokens on the board, Sarah was rewarded with both physical and social rewards. For example, in order to receive an apple, Sarah was required to place the two tokens for *give* and *apple* in a vertical sequence on the board. Although not without difficulties, the advantage of this system over the gestural signing languages is that as a sequence of instructions is given to the animal in the form of ordered plastic symbols in a particular serial order, the internal retention of the sequence is not crucial since the

external record remains visible. By thus reducing the requirement of memory for the manipulation of signs, or sequences of informative gestural symbols, Premack's task has become a less demanding one for the subject. However, not only was this new system to remove the pressure of memory from the problem-solving situation, it was also to remove the pressure from the subject in its being required to choose its response from large set sizes of alternative responses for each of the problems posed. Although early results with Sarah lead Premack (1976) to claim that the chimpanzee was capable of classification (X is a ?), analogous reasoning (A is to B as C is to ?), propositionality (including conservation of volume) and cause and effect constructions (whole apple + ? => cut apple [? = knife]), nearly all of the target responses were chosen from binary pair choices made available. Furthermore, all of Premack's sorting experiments took place with regard to unordered items (there was no requirement to disorganise spaces) and the chimpanzees were never given pre-organised items to determine whether this might reduce the disposition to either organise or reorganise the stimuli sets made available.

In summary, despite Premack's chimpanzees demonstrating a lexigram vocabulary of more than 130 symbols (in the case of Sarah), he was unable to provide any evidence for the animal's production of structural (lexical) innovation. Successful substitution was clearly evident for a variety of object, events and simple requests for information (and action) to be provided, but nowhere in his writing does he offer an account of instances of the chimpanzee providing any kind of grammatical classification or otherwise engagement in either relativisation or nominalisation. It is quite fair to conclude, therefore, that the performance of these chimpanzees (as with those of the Gardeners before them) does not differ that much from the performance shown by the parrot (Alex) or indeed by the pigeon who has learned to peck one response key rather than another in order to receive access to food. One might object to this view in proposing that associative learning plays an important part in the acquisition of a vocabulary in the first place, but, in order to support the claim that the chimpanzees had acquired a language (as the researchers so far discussed above in this section do) one might reasonably expect a little more sophistication on the part of the animal's performance. In particular, their case would be more convincing if the chimpanzee could be seen to use the words that they had learned in grammatical sentences,

in combinations and recombinations of words appearing in appropriately ordered structures. Such evidence remains unavailable from the symbol-based language research to date. One final avenue of research using this approach with the chimpanzee, but with quite a different interface apparatus, was developed by Rumbaugh (1977) in the LANA project.

Using a computer-controlled lexigram console with a purpose built symbol keyboard system, a single animal could exchange information without the need for the experimenter's presence during the interaction. The advantage of this novel situation was that symbolic exchange with the computer equipment could occur without the possibility of the human trainers providing external cues during the interaction. Unlike the situations found with the signing experiments, in which trainer imitation commonly appeared (Terrace et. al., 1979), the use of this new lexigram keyboard required that the chimpanzee exhibit its correct responses only by working them out for itself, without being able to assess subtle changes in a trainer's posture, tone of voice or facial expression. Although the early training was rather difficult, the chimpanzee (Lana) came to reliably use a vocabulary of some 70-80 of these visual patterns used as symbols ('words') to denote objects (e.g., foodstuffs, toys and simple tools). Intensive individual training of Lana established that she could produce a wide range of requests for desired objects and events (a four lexigram sequence known as *Yerkish*, translated, for example, as '*please machine give juice* period'), and that relatively clear evidence was provided for references being made to her 'desire' for absent objects as well as visible (but unobtainable) foodstuffs, toys and simple tools (Rumbaugh, 1977). However, little inovation was forthcoming beyond their trained sequences and, as a review by McGonigle (1980) rightly points out:

> 'for example, the words *please* and *period* are fixed features of the sentence frames ..... [and that] the analysis did not include any consideration of the communicative or the contextual appropriateness of these grammatical strings and it seems likely that not all of the constituent lexigrams in the strings carried an independent meaning for Lana'
> (McGonigle, 1980: p. 286)

In a later step, again using specialised training procedures, two chimpanzees, Sherman and Austin, were used in a 'communication experiment' in which the two animals were required to cooperate with each other with their respective keyboards (Savage-Rumbaugh et. al.,

1978). In a typical experiment, one of the chimpanzees (only) would be presented with a situation in which a foodstuff was placed nearby within a container requiring manipulation with some tool prior to its being opened. A second chimpanzee could see the first, and his lexigram keyboard output, but neither the foodstuff nor the container was visible. What was also available to the second animal, however, was a set of tools from which items could be selected and passed to the first via an adjoining 'cat-flap' between them. Only by providing the first chimpanzee with the appropriate tool to solve his problem via the observation of lexigram requests, the two cooperating 'communicators' were subsequently able to share the previously 'hidden' food rewards with each other. The results indicated that the previously learned arbitrary patterns of symbols had come to elicit representations of the foods and tools that they signified, and that each could act as either sender or receiver with a success rate of 95%. The researchers may rightly claim that these performances have 'constituted the first documented instances of symbolically-mediated exchange of goods and information in a non-human species' (Savage-Rumbaugh et. al., 1978, p. 540), but even these dual-coding experiments have not given rise to the exchange of any new information suggestive of the chimpanzee's wishes or intention to do anything other than to obtain either food or tools as determined by their training schedules. However, some researchers have gone further with their claims, and, using the results of other 'hidden food' and communication experiments with non-human primates, are willing to support the case for the animal's possession of a 'theory of mind'.

An animal 'theory of mind' ?

Following from the work of Menzel & Halperin (1975) with chimpanzees kept in open-field enclosures, much interest was developing in the idea that troop leaders might be aware of the communicative value of their signals. In a typical experiment, the whole troop would first be brought to a caged enclosure out of sight of the main open-field area. One member of the group (the 'leader') would then be taken, alone, to a site outside and shown a hidden food cache location before being returned to the troop. Sometimes a second chimpanzee would also be taken separately to view the location of a hidden toy. After some 10 minutes post reunion, all of the troop was once more released into the open-field enclosure and their

individual activity monitored. It was discovered that most of the troop first visited the hidden food location prior to examining the hidden toy, and, as only one animal knew the food location it was suggested (Menzel & Halperin, 1975) that the signals transmitted by the 'leaders' must have informed the others of the relative value of the hidden objects and where they might be found. Although some commentators interpreted these results as providing 'significant evidence of conscious intention on the part of animals capable of moderately complex communication' (Griffin, 1978, p. 534), this is by no means an obvious solution. But quoting from the original reports:

> 'Usually each leader took a few steps toward the goal object that we had shown him (or her) earlier, and then stopped and looked back at the rest of the group. If the other leader was setting out or trying to recruit followers more vigorously than he (which usually occurred if the goal was a more highly preferred one), he often abandoned his own goal, accompanied the other leader, and then led the group to the second goal.'
>
> Menzel & Halperin (1975: p. 654)

No evidence was ever found for intentional variations in the types of detectable signals emitted as a function of whether hidden objects might be foodstuffs or toys, and nothing more than the vigour with which a 'leader' moved towards a goal needed to be observed in order to determine what the goal might be. Even the claims of Whiten and Byrne (1988, 1997) for the 'tactical deception' of 'crafty chimpanzees' suppressing cue-signalling upon their discovery of hidden foodstuffs (cf: de Waal, 1989) may be explained in terms of the behaviour of one animal serving merely as a stimulus eliciting certain predictable responses from the other. Another variation of the 'hidden food' problem, in which a chimpanzee indicated the location of a food to a naive human partner (or vice-versa) was devised by Woodruff and Premack (1979).

Following from their own work with video film and 'what happened next' photographic stimulus experiments (Premack & Woodruff, 1978) in which the chimpanzee was said to have understood the goals of a human actor in the film, using the same animals, the subjects were later shown to be able to differentially respond to the particular 'intentions' of an actor. Now using 'co-operative' and 'competitive' participant actors who offered either helpful or misleading advise respectively, pointing gestures and

postural cues aimed towards or away from hidden food locations in the chimpanzees environment were successfully responded to after some experience with the task. As with the results of other 'direction of gaze' studies (Povinelli & Eddy, 1996; Whiten, 1997), these experiments are rather difficult to disentangle, but they do nonetheless provide some support for the contention that postures and gestures may be given varying interpretations by chimpanzees dependant upon the individual in whom they are observed (see also various contributions to Whiten & Byrne (Eds), 1988, 1997 concerning 'social intelligence'). Similar findings have been derived from audio-playback studies of the alarm calls of the vervet monkey (Cheney & Seyfarth, 1990) in which the recognition and adaptation to the calls of specific individuals has led to differential responding according to the listener's experiences of that individual's call frequency and reliability in terms of real threat indication. Cited by Griffin (1984) as again providing 'confirmation' of the existence of non-human animal thought, these findings may be simply explained by associative mechanisms of action without the need to require some notion of conscious awareness, 'theory of mind', or other great intellectual dexterity on the part of the monkey. At the very least, if one is to be recruited to Griffin's view that 'a possible window on the minds of animals' is available following 'participatory investigation of animal communication', some explicit test of subjective awareness (which is not necessarily the same thing as cognitive complexity) needs be proposed. Indeed, neither in his revised (1981) edition of *The Question of Animal Awareness:: Evolutionary Continuity of Mental Experience*, nor in his subsequent works (Griffin, 1984, 1992) does he offer any theory of cognition as such, remaining content merely to argue that the behaviour of animals should be discussed in terms of their awareness of mental events, irrespective of degrees of abstraction or the cognitive complexity that may be inferred across a variety of species. With respect to this latter issue, it is a further weakness of the case for an animal's 'Theory of Mind' that Griffin and his supporters would appear to give equal weight to the gestures of chimpanzees and the signals exchanged by honeybees, as prototypes for human language.

Lessons to be learnt from the ape-language studies

The importance of the results of Premack (1970, 1971, 1976), Rumbaugh (1977) and Savage-Rumbaugh et. al., (1978), taken together with the other work with apes reviewed above, is not that the apes can be trained to do things that might plausibly be equivalent to human language. Rather, it is that at very best, apes have been shown to be capable of being trained to do things which might plausibly be equivalent to some preliminary stage in the evolution of human language, or which indicate that apes might possess a level of cognitive organisation which one may imagine could be sufficient to make the beginnings of speech useful. Indeed, in asking what had been revealed about the mentation of apes following their 'language tutoring' that was not inferable from other aspects of their behaviour, McGonigle (1980) also argues that:

> 'On the one hand the content of ape language merely reaffirms the narrow incentive range within which the chimp seems to operate in laboratory situations.'
> 'On the other hand, the restrictions on semantic role .... are symptomatic perhaps of an egocentric form of thought - a characteristic of early stages of child development according to Piaget - not easily revealed by non-verbal tests.'
>
> (McGonigle, 1980: p. 286)

Although Premack (1976) has claimed that a 'language facility' makes animals smarter by allowing them to 'think' in terms of abstract symbols rather than they would with visual images only, there is little evidence to suggest that the 'language-trained' chimpanzees have solved real-world problems any quicker than have their untrained peers. Furthermore, the chimpanzee's use of abstract 'words' and symbolic thinking skills as so far reported would not appear to readily lend themselves to any degree of *extendibility* beyond their providing simple tools for the chimpanzee in situations specifically designed by the researchers to make such 'word'-use necessary.

The results of these 'ape-language' studies to date have served to confirm the earlier suggestions that were based only upon less sophisticated behavioural experiments (such as those reported by Köhler (1925), but they do not really contribute much to one's clearer appreciation of the intellectual powers of other species. What still remains lacking therefore,

is the production of any coherent set of procedures embedded within a single paradigm with which one might attempt to bridge the apparent impasse brought about by the appearance of human speech and language, a behaviour indeed unique to man amongst all of the animals. One proposal from McGonigle (1980) is that:

> 'The responses of apes to attempts to teach them the spacio-temporal conventions for reference to objects and events outside the immediate context of utterance should provide further important clues to the cognitive resources available'
>
> (McGonigle, 1980: p. 286)

The road ahead, therefore, will necessarily be paved with a number of minimal requirements prior to our seeing a more revealing account of the development of comparative animal cognition and any progress towards our better understanding the continuities between the mentations of human and non-human animals.

Prospects for a developing comparative psychology

It has now become clearer that the procedures and paradigms to be implemented and developed in the future must consider, and incorporate into their design, a number of identifiable features. The failings of both Spence and Darwin, although not lacking in theoretical consideration, lay in their reliance upon the accumulation of largely anecdotal data sets and want of empirical studies. Romanes and his followers were likewise to only infer mentation from behavioural observation alone and to be able to deduce the existence of ideational thought only from introspection and attribution. The continued failure of these introspective methods were to give way to the rise of Behaviourism, a movement promoting a stricter methodology. But although grounded in more scientific methods (sympathetic to Pavlov's anti-mentalist position) Behaviourism was to confuse its methodology with the objects of comparative cognition, the study of behaviour itself becoming the direct end of the psychological investigation, rather than it being the indirect means of accessing any extant mental processes. However, mainly through the efforts of Skinner, a number of important contributions were to be gained from the Behaviourist approach including the technique of 'constructing' an experimental animal's behaviour for subsequent manipulation in

controlled conditions (rather than using behaviour trouve - see also McGonigle and Chalmers, 1996a); the shaping of voluntary behaviours (rather than of the merely autonomic functions seen in Pavlov's laboratory); and the repeated use of individual animals over extended periods of time so as to allow their ontological profiling during their individual life histories (rather than using short, snap-shot studies involving different individual animal groups for comparison). Further to these concerns, in order to prevent the all too frequent occurrence of the 'levelling-effects' found in some of the animal learning experiments discussed above, it would appear to be necessary to employ tasks which are by their nature extendible and which may be altered in such a way as to require the subject to continually adapt to subsequent pressure (i.e., to continue to monitor and regulate its successful behaviour according to increasingly stressful task demands), whilst not readily appealling to associative interpretations of their behaviour (see discussion of Harlow above). The call then, is for the development of procedures and paradigms which afford the assessment of the degrees to which a given animal can demonstrate successive adaptation of its behavioural repertoire in the face of tasks with increasing levels of difficulty. Furthermore, each animal's successes (or failures) with such tasks should be monitored (by both agent and experimenter) over protracted periods of time (years rather than months) in the course of its individual, life-historical development.

The weak induction hypothesis: developing a new typology and paradigms for developmental and comparative animal cognition

In hindsight, perhaps the conclusions culminating in the statement of Macphail's (1982) 'Null Hypothesis' are not altogether to be unexpected. But the types of learning which have traditionally been investigated and reviewed above may not be the only ones. Whilst adherents of behaviourist and associationist's views have sought to confirm their theories using learning by association tasks, blind trial and error types of learning may be the most primitive and the worst sorts of learning to provoke species differences. Instead, better differentiation may be expected using other forms of learning, such as the insight-based types Köhler (1925) had argued for many years ago. The one area of investigation that had revealed apparent species differences, whilst at the same time producing results not wholly accounted for by associative

mechanisms, was that in the use of techniques developed by Harlow (1949). The crux of the matter now, however, is to establish an alternative typology which, whilst recognising the adaptive role of associative learning, also specifies new forms of learning-based adaptation from which a more fecund comparative psychology of learning could derive. Addressing this issue, McGonigle and Chalmers argue for (at least) two types of learning mechanism, firstly a weak, and old association-based one, and secondly, a stronger relational based mechanism.

As these authors (1996) point out, association-based mechanisms are weak inductive devices which require a very reduced space of induction bounded by the small time and space intervals between events necessary to induce conventional trial-and-error based learning. Characterised as an evolutionary 'quick fix' which presupposes only limited resources on the part of an agent to operate such a system, the procedures involved, which keep the spatial and temporal interval between discriminandum and manipulandum to a minimum have the consequence that:

> 'few irrelevant events can distract from contingency based solely on the very short (optimal) time and space intervals involved in connecting stimulus and stimulus, and stimulus and response'.
>
> (McGonigle and Chalmers, 1996: p. 2)

As for alternatives, the authors propose one, qualitatively different, and based on relational mechanisms which have had no place historically within association-based learning paradigms. As a result:

> 'Whilst the mechanism reflected in these [association] paradigms are "old" in evolutionary terms, they do not cater for the significant distinction between arbitrary and non-arbitrary connections between objects, nor do they (nor can they) specify how evolutionary systems invest in complexity, and become more powerful both as engineered by evolution and as a consequence of their own individual life histories.'
>
> McGonigle & Chalmers (1996, p. 18)

The position taken here is not, however, one of wishing to reinterpret the findings of the associationists' experimental findings in order to make claims for the existence of higher cognitive processes. It is rather to point out that the constraints quite necessarily being placed upon an experimental subject during the employment of stimulus-response, three-

way contingency methodology, do not provoke a rich enough depth of analysis in the subject so as to illicit the demonstration of adaptive intelligence to any significant degree in the first place. However, unlike Karmiloff-Smith (1991) for example, McGonigle and Chalmers do not hold the view that evolution and development involves a serial unfolding of progressively abstract layers of cognitive competence, wherein each later stage of competence is a derivation of the one immediately prior to it. Indeed, they offer the proposition that:

> '... in complex systems at least, a wide spectrum of high level cognitive behaviours have hard-wired underpinnings as *necessary preconditions* for their (ultimate) expression in later cognitive development.'
>
> Italics preserved, McGonigle & Chalmers (1996: p. 19)

If McGonigle and Chalmers are correct, for both the evolutionary and comparative questions to therefore be addressed, a different methodology, or paradigm shift, would be required in order to demonstrate the existence of any extant higher cognitive behaviours. If their claim for the coexistence of different sorts of adaptive competences as lower bound design primitives is to be sustainable, then, as expressed in their own words: 'the less rich and diverse the primitives, the more limited the system ontogenetically' (McGonigle & Chalmers, 1996: p. 19). The problem of exactly how to design and implement such a new paradigm which takes account of these shortcomings, is the subject of the next section and subsequent chapters. This new paradigm aims to produce longitudinal, comparative and developmental assays for analysis, using tasks which will elicit a sufficient depth of cognitive processing whilst keeping individual subjects (both human and non-human) engaged and challenged. Unlike the procedures employed in the research reviewed above, rich behaviour graphs over relatively long periods (e.g., years, for bird and monkeys) are undertaken, and the levels of task difficulty be made variable according to the individual subject's success profile over its own history throughout the experimental series. With such new procedures, it is to be hoped that one might see exactly to what extent the subjects might epigenetically regulate their behaviour, using their own success in earlier tasks experiences in order to solve later problems of new levels of difficulty, previously too difficult for them to entertain cold. Such procedures, if workable with human and non-human species, will allow new measures to

be made together with an evaluation of McGonigle & Chalmers' (1996) claim for an 'agents growth as inductive system, from weak to strong'. The genesis of this new paradigm saw its roots in the work of Harlow (1949) on learning sets and, although it remained poorly characterised, will be reviewed here so as to enable a better understanding of its extendibility to include a hierarchy of task-level difficulty (McGonigle & Chalmers, 1996, 1997) and to assess the support for a weak induction hypothesis for the development of animal learning.

Biological implications of learning

What the psychologists now had at their disposal was a powerful tool in the form of the learning experiment which came to provide a frame of reference enabling them to explore issues in comparative cognition in a way not previously open to the ethologists (Dickinson, 1985, 1980; Hendriks-Jenson, 1995). The problem for the ethologists was that for each of their cognitive accounts of animal behaviour, there was an equally plausible counter-argument from the behaviourists. Not being sufficiently grounded, the theoretical position of the naturalists failed to specify what might happen according to the different theories due to their lack of experimental paradigms and procedures for empirical evaluation. The achievements of both Lamark and Darwin, in their day, was that both provided acceptable *a posteriori* explanations of the causes of adaptations. However, after Lamark's explanation later had appeared to be wrong and Darwin had come to be considered at least generally correct, it was for the later neo-Darwinists (see Gould, 1977, for review) to point out that adaptation to local conditions was only a primary cause of the morphological and behavioural changes seen amongst the great diversity of species. Recognising at least two other sources of change in an animal's ontological development, the Darwinists spoke of "correlations of growth" (allometry) and exadaptation (Gould, 1977). For the case of allometry (the systematic study of animals' growth and size) such correlations between the relative sizes of, say, brain and body are not simply explained and contain many anomalies. But Darwin (1871) himself had also entertained the possibility that structures or behaviours might occur which supported functions different from those for which they might have originally been selected. Having thus been recruited to some such 'new' function, although it may not constitute the optimal functional solution, it thereby

provided a solution nonetheless. Only later named 'exaption' (originating either as a non-adaptive feature or first evolved for some other use) such an adaptation might depart considerably from what an engineer might regard to be an optimal design. Although most adaptations will result from a compromise between some perfect functional solution and what can in fact be achieved, as pointed out by Hendriks-Janson (1995) in his discussion of Mataric's robotics work, '..... emergent functionality is always situated'. The point being made here by Hendriks-Janson was that emergent functionality and preadaptation (be it cognitive or not) will appear similar due to their desired design outcomes being equated to evolutionarily selective pressures.

The development of intelligence as adaptation

A continuing criticism of adaptational theories (owing largely to Lewontin, 1981) is that too static a view is often taken of the idea of adaptiveness; that it is seen as the solution of a passive creature to the dominating forces of its environment. It requires little thought, however, to realise that all of the great diversity of living creatures change the world in which they live and grow in a variety of ways. Not only are man and the tool-using great apes involved in this exchange, even an oak tree casts shadows, sheds leaf litter, removes nutrients and water from out of the ground and releases oxygen into the air. Trees, plants and all animals do not, therefore, just *react to* their environment, they change it too, and so are in dynamic interaction with it. In the case of the primates, their high degree of mobility and manual dexterity will, if anything, increase the dynamics of such interactions as they might have. In this sense, adaptations are not merely static solutions to static problems and thus, if the evolution and development of animal cognition is to proceed via this process (and there is no reason to suggest that they will not), then any program investigating the nature of animal intelligence cannot assume the existence of appropriately developed cognitive processes. Throughout the first half of this century, the studies of by Köhler, Yerkes, Harlow, and later the behaviourists, had seen progress towards a more empirical experimental psychology. At the some time, however, they did remain content to assess the existence (or non-existence) of assumed cognitive processes via snapshot studies, at best providing existence-proofs of target behaviours using short-term learning experiments. What all of these

studies had in common, was the persistent lack of warrant being given to the notion of a developing theory of task on the part of the experimental animal undergoing examination. More importantly perhaps, little consideration was given to the degree of expertise to be expected from the subject as a result of its ontologically developing life history. This latter point cannot be stressed enough, as the degree to which an individual subject might be expected to demonstrate success with any particular type of task will necessarily be the result of that subject's previous interaction with both similar and the actual materials and details of the task environment. This neglect of life-historical factors interacting (or indeed, in certain circumstances, determining) the ways in which any subject might be expected to report success in a learning experiment does not take into account the extent to which the subject might be task-informed. This is of great importance when attempting to interpret a subject's failure as much as success at any given task. Unless the study of animal learning and cognitive development be afforded some special mechanism of operation (and there is no reason to propose that they should, at present) the dynamic nature of adaptations are going to be as characteristic of a particular set of adaptations called intelligence as they are of those concerned with, say, respiration or muscle tone. An individual animal's competence will be determined by their own individual adaptations, some of which are continually interacting with the experimental environment and history of task success. However, although it is true to say that selective pressures (whether created by either novel environmental or experimental situations) will always favour the preadapted animal, one cannot simply equate intelligence with adaptiveness. As proposed by McFarland & Bösser:

> 'Intelligent behaviour is not the product of a particular mechanism for generating behaviour. Although intelligence is often equated with cognitive ability, cognitive mechanisms may be inferior to automatic hard-wired mechanisms in certain ecological circumstances'.
>
> (McFarland & Bösser, 1993: p.281)

The difficulty one is thus faced with, is to imagine a mechanism whereby a stable (or at least, a successful) organism may come about which has a degree of continuing flexibility concerning its behavioural repertoire in the face of familiar and unfamiliar stimuli, and yet at the same time maintain any previously 'acquired' knowledge (derived from either

ontogenetic or phylogenetic processes) which continue to serve the satisfaction of its basal metabolic needs.

Dynamic interactionism and epigenesis

In the 1960s Waddington argued that evolution required both a stable memory store and what he called on 'operator' component (Waddington, 1969). The operator was not another passive-reactive entity as championed by the behaviourist Watson, but was here invested with dynamic, reactive qualities, one of which was the capability in some cases of altering its own environment. Such operators were what made life 'interesting' according to Waddington, whereas the memory stores were rather dull and unreactive - (DNA, indeed, is an exceptionally inert chemical). As suggested above, whereas trees are relatively immobile, essentially only making use of the nutrients close by them, the case of animal lifestyles and especially that of the mammals, has necessitated specialist adaptations for motile living. The requirement for movement, and the associated consequences of that movement, have imposed very fundamental selection pressures upon the evolution of organs which can co-ordinate such movement, and ultimately for the evolution of learning and intelligence (sessile organisms are essentially devoid of any clearly identifiable central nervous system (Greenfield, 1997)).

Such an adaptation involves not just internal-external relationships but internal-internal ones as well, and such inward looking adaptive relationships may be of special significance when one is considering the adaptive states of brain structures and their concomitant behavioural adaptations (Edelman, 1986; Changeux, 1994). However, whether involving inanimate objects, other organisms, or closer conspecifics, adaptations are never static things but dynamic in their interactions with the world. In the same way, the development of an animal intelligence must surely be also viewed as dynamic and interactive. The details concerning how this process might take place remain illusive at present, but it is pertinent here to offer a brief discussion as to what evolutionary pressures might lead to the development of animal learning and intelligence as part of the same mechanism provoking the development of the more apparent behavioural and gross physiological developments readily observable in the field.

The problem of instinct rules and predictability

By the employ of similar reasoning to that applied above to the case of increasing motivation for the control of mobility, the requirements for successful learning in a given individual animal, and the associated consequences of that learning, will impose fundamental selection pressures upon the evolution of physiological and cogniitive processes which underlay such learning behaviour. Although both Skinner and Romanes long before him had alluded to this notion, neither of them had incorporated the opportunity for an animal to demonstrate such an effect into an experimental program which might have revealed the growth of 'higher' mental processes in the laboratory animal. Views as to the extent of any adaptive behaviour across species vary, but many authors contend that at least five per cent of extant species supplement their instincts with additional devices for shaping adaptive behaviours (see, for example, Plotkin, 1988). Likewise the time course of events between receiving a set of genetic instructions and their implementation, through development to the point where those same instructions might be returned, via reproduction, to the gene pool, has been variously described (Dawkins, 1989; Lorenz, 1971). In an unchanging world this is not necessarily a problem. But in effect, what nature does is to constrain organisms to act within a future about which there can be no certainty, providing a set of instructions for the construction of adaptations, some of which may no longer be appropriate for the successful ontological growth of an organism in its new environment. Indeed, as stated by Waddington:

> 'The systematic exploration of the evolutionary strategies in facing an unknown, but not wholly unforcastable, future..... take(s) us into the realm of thought, which is the most challenging..... of the most basic problems of biology. The main issue of evolution is how populations deal with unknown futures.'
>
> Waddington (1969: p.278)

What is being addressed here, is the question of how life maintains itself on the basis of inadequate information. The genetic-developmental logic models of evolutionary theory have two principle features (McGonigle, 1991; Maynard-Smith 1982). One is that it takes the logical form of induction, generalising into the future what previously had worked in the

past (a conservative, pragmatic component). The other involves the generation of novel variants by chance processes (the radical, inventive component). Together these are nature's way of injecting new variants into the system in order to (potentially) make up for deficiencies that may occur if what worked in the past no longer does so due to both the organism and/or its world having changed over time. It is precisely because such changes are sudden, effectively without precedent, that the inductive logic of the conservative component cannot supply effective adaptations to cope with them. Only if there is an element of repetition, of predictability, of projection into the future of what occurred in the past, can the habits-of-mind or hard-wired processes work inductively as a source of adaptations. Unpredictable change is a form of Waddington's uncertain futures problem as mentioned above, a problem that cannot be solved by a purely inductive process.

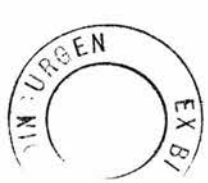
The growth of cognitive processes as adaptation

The message being delivered here from evolutionary biology to the psychologists, is that perhaps the levelling effects so commonly derived from animal learning experiments in the past, might be partly explainable in terms of the methods and procedures employed. Using mostly static, snap-shot experimental designs, one outstanding feature of the traditional behaviourist procedures was that the animal was rarely put under any continuous pressure to resolve problems of an increasingly complex nature. From the viewpoint of Waddington's evolutionary epistemology, unless hard-wired in advance, an animal would only be expected to demonstrate behaviour indicative of 'higher' cognitive processes if exposed to a series of ongoing test situations which required the production of continuous and appropriately made adaptational responses. It is to this last issue that return will be made in the next chapter, when laying bare the rationale for the experimental work to be implemented and reported in this thesis. The specific hypothesis to be tested (e.g., McGonigle and Chalmers, 1992) relates directly to this biologically informed epigenetic attitude, for which one might propose that both the role of cognition in development, and the specific regulators of cognitive growth in a particular individual, be dependant upon the successful management of progressively taxing environmental problems in need of a solution. By this reasoning, it is only by providing an

environment in which a test animal is subjected to progressively demanding and complex tasks, that one might expect to see the emergence of any 'higher' levels of cognitive organisation which might be sustainable.

Earlier in the chapter, it was seen that the behaviourists (e.g., Skinner, 1977) had overemphasised the control of behaviour and thereby diverted attention away from complex problem solving conditions in favour of artificially oversimplified task environments. This situation in turn, was responsible for a major confusion of the methodology and the objects of comparative cognition. The observable behaviour became the direct end of psychological investigation, rather than the indirect means of accessing mental processes, and it could be seen that the evolutionary framework which originally constituted the principal rationale for comparative psychology thereby lost (or, at least changed) its meaning. On the very best analysis, differences in the complexity of various species' behavioural repertoires were accounted for by simple summation functions of associations which had been more or less rapidly established (see also McGonigle, 1991). At the same time, the failure of non-human species to convincingly exhibit 'higher cognitive' functioning in the laboratory, and in the finding of the 'levelling' effect of their learning results, were said by many (e.g., Macphail, 1982) to be indicative of the absence of a linguistic competence possessed only by the human subject.

Our journey through the last century or so of comparative psychology has shown that although the natural historians and ethologists had conducted little empirical work, some progress was later to have been made by the more experimental psychologists. We have seen that the kinds of problems previously given to laboratory subjects in typically short term, snap-shot experiments were not of a type which might necessarily demand of the subject a more cognitive solution. And neither did they make address to the epigenetic growth position as was then developing in evolutionary biology, which would suggest the use of a task situation that could not be solved by purely inductive processes alone. This latter argument has important implications for any interpretations laid upon the results of the animal learning studies conducted right up until the present. What was necessarily required, and until now had been almost entirely absent, was for any study exploring the existence and development of animal



cognitive processes to have taken place over long periods of time, using the same subject, whose levels of competence and behavioural adaptation could be continuously monitored throughout the duration of a more longitudinal study as it progressed.

Chapter 2

Towards the need for a new paradigm in comparative animal cognition.

If the indications from Harlow's animal work were correct, and cognition required on-line judgements to be made concerning the most likely consequences of future behaviour, then such an evaluation would necessarily imply a high degree of motivational autonomy on the part of the subject. Evidence for such a claim in the non-human animal literature certainly did not come from the behaviourist camp (heavily investing in merely behavioural modification rather than the construction of behaviours for subsequent adaptation), but neither was it likely to derive solely from the 'learning set' type of methodology which also used relatively short-term snapshot studies with different animal subjects merely solving an endless series of binary discrimination tasks. What was needed in order to demonstrate a greater degree of autonomy, was a paradigm within which an extended choice of behaviour could be seen to be drawn from a definable set of possibilities, using the same subject over a protracted period of time (years rather than months). Furthermore, in accordance with Waddington (1969) the same subject needed to be exposed to an ongoing series of tasks of increasing levels of difficulty throughout the course of the study. Only by satisfying these minimal requirements would any animal subject be provided the opportunity to undergo, let alone demonstrate, any cognitive growth and development such as they might experience over their individual life span. Although an animal capable of complex intelligent behaviour would otherwise need a relatively large brain (because all responses to possible states of the world would have to be pre-programmed) this problem would be eased by following the development of an animal with at least a partial autonomy throughout its own lifetime, plus the addition of some 'cognitive abilities' (so allowing evaluation of trade-offs or some degree of optimisation). How then, as a first solution, might such an organism be 'designed', or prepared, so as to be able to cope readily with predictable changes ?

The need for longer term studies

It is within any one life cycle that changes of salient importance take place, and a local observer, be it a human, monkey, or pigeon, experiences the cycle (i.e., one's life) as containing such changes. If the life span of the localised observer is less than the period in which a cycle of changes has taken place, then that organism cannot possibly acquire information concerning the changes that make up the sub-cycles occurring throughout its own experience. Likewise, if a learning experiment is set up in such a way that the subject is not sufficiently task-informed or in some way 'tuned into' the salient features of the stimulus array to be interacted with, over a sufficiently long period of time, then its depth of exposure and analysis will remain shallow. Likewise, it will not be susceptible to any adaptive responding as might conceivably occur over many cycles following the detection of such features, be they natural variables or changing experimental parameters. Even worse for the investigator conducting animal learning experiments concurrent with the more invasive neuropharmacological or terminal neurophysiological techniques, the subject is often no longer intact and available for further enquiry after the tasks have been completed, let alone qualify the same animal for subsequent investigation. The problem for such methods in the study of animal cognition, is that they disable any attempt to follow an individual animal's cognitive growth trajectory from its initial learning experiences. For these particular types of subject pools, one could never come to know what later intellectual developments might have emerged had the animal survived to continue its learning program (in the light of its previous learning experiences). Indeed, without a plotted history of the individual subject's successes and failures with any given task, the issue of an animal's 'learning to learn' throughout its long-term maturation would not be an option for investigation using the traditional methods of neuropsychology. Such adaptations which might have revealed themselves, may typically require longer periods of time to establish themselves than are normally allocated to these kinds of experiments. Furthermore, with the current techniques of developmental histology, even when attempts are made to pursue a longitudinal study, the nervous tissue of the animal is often disrupted in ways inimical to the replication of a more naturalistic scaffolding of cognitive growth over time. Additional to these issues for the establishment of conditions conducive to

sustaining experiments over a number of years, are considerations concerning the maintenance of the subject's vitality and social welfare.

The role of health and social welfare in determining the cognitive growth trajectory of the experimental laboratory animal

In order to facilitate and sustain the health and welfare of a non-human laboratory subject, one proposal might be to provide a higher standard of husbandry care than is normally seen in animal laboratories. Reference is not being made here just to the social withdrawal and stereotypy problems frequently reported to result from the use of isolated, single caged animal housing so commonly in the past (Singer, 1991; Stamp-Dawkins, 1992). The principal concern here, is that the laboratory conditions (including the animal housing, handling, feeding schedules, social and family grouping interactions, etc..) should not be divorced from the details of the animal subject's more controlled experimental session arrangements. As discussed above, the situation in the natural world for any given species is such that their habitat conditions (and fluctuations) both support and interact with the animal's physiological development. It is therefore of the greatest import, in order to assess the potential growth towards some degree of cognitive expertise in, say, a monkey living in a captive, managed, animal laboratory, that one simultaneously allows the animal's more natural growth curve to progress in as rich and supportive social environment as can be managed, whilst at the same time catering for the physical well-being of each subject. Such conditions are essential to the ongoing evaluation and assessment of both the successes and failures during an individual animal's progress, especially when monitired throughout the duration of a long-term study lasting a number of years. Especially in the case of task failure interpretation, this is of crucial importance for two reasons: firstly, the literature to date does not offer much indication as to what might be the expected levels of expertise that might be already in place at a given stage of development, so requiring the careful design of the earlier achievable tasks for evaluation to take place on an individual basis. Secondly, the specific design of experiments for which the subject might suddenly fail in the face of rising task difficulty, can be better assessed if such failures occur in the absence of other confounding social and welfare factors that could otherwise affect their performance.

In summary then, in order to better evaluate the findings of a long-term study, using the same subject over a protracted period of time, it is essential that the experimenters have integrated the animal's developing growth trajectory and any increasing expertise, with the development of laboratory conditions for the purposes of undertaking a longitudinal study. This has very rarely been the case, and more typically in the past, for example, many laboratories have employed subjects drawn from a bank of animals for short-term experiments, kept in isolated caging with little or no social interactions in an animal house within which psychopathies were common. In order to allow for the better evaluation of the results of investigations in comparative animal cognition, especially in the face of failure at a task, the experimental environment must be designed (and allowed to continuously evolve) so as to provide the most optimal conditions conducive to the production of a developmental profile as near as possible matching that of the animal's more natural disposition in the wild. Given that the provision of an optimal environment for the study of comparative animal cognition can be met with respect to the needs outlined above for a longitudinal program of studies, what the particular behaviour-based conditions might be which could enable a staged growth of cognitive expertise to be revealed remains to discussed and determined. Likewise, for the study to be reported below, how might we lead, say, a monkey (a non-linguistic primate subject), through a series of informative baseline experimental conditions, prior to subsequently increasing the levels of difficulty as determined by the animal's ongoing task successes ? Rather than attempting once more to learn from the mistakes and shortcomings of the ape-language studies or the animal 'theory-of-mind' literature once more, we will now take a brief look at some of the claims for the development of *human* cognition and language in an attempt to determine some clearer direction.

The development of human cognition and language

Despite the improved recommendations for a more empirical approach to the growth of cognitive behaviour in animals, there still remains the need to address the means by which we might determine the extent of any judgement criteria alreadily 'built-in' to an evolving organism such that it be able survive its formative learning stages. This is really no lesser task than is required prior to our exceptance of the claims for sophisticated

human adult cognition (which is so often explained away as resulting from the human's possession *per se* of language) and communication. Moreover, whilst skilled adult behaviour appears both directed and flexible at the same time, views of development such as those of Karmiloff-Smith, (1992, 1993) do not go on to propose any putative mechanisms which might enable this to take place. No matter how small the "atoms of action" that might result from "representational redescription"(see further discussion below), the fact that they are prespecified and are explicitly combined into an overall plan prior to activity imposes constraints on the behavioural repertoire to be exhibited. What is required is a view of development which imposes few, if any, limits on the potential flexibility and context sensitivity of behaviour whilst also allowing deliberate shaping to take place. Further, categories appropriate for the encoding of information for the purposes of communication are not likely to be the same ones as those used for the internal control of behaviour. Behavioural control cannot be goal-directed in that sense, although it may provide for a useful way of representing behaviour for the purposes of communication. In a similar vein, there has, to date, been no clearer characterisation of the notion of 'stimulus control' (as used by Skinner) that was usefully related to his bar-pressing experiments. Indeed, as Chomsky (1959) points out:

> 'What has been hoped for from the psychologist is some indication of how the causal and informal description of everyday behaviour in the popular vocabulary can be explained or clarified in terms of notions developed in careful experiment and observation, or perhaps replaced by a better scheme. A mere terminological revision, in which a term borrowed from the laboratory is used with the full vagueness of the ordinary vocabulary, is of no conceivable interest'
>
> (Chomsky, 1959, p558)

As with our treatment of Karmiloff-Smith (see above, and next section below), what Chomsky was doing there, was to accuse Skinner of redescribing, rather than explaining, verbal behaviour. Chomsky (1959) ends his paper by making a strong case for competence theories, in which complex behaviours could not be explained as deriving from a succession of reflexes. He suggested that complex behaviour (like language) had a syntactic structure that must be imposed from within, and referred to another classic paper by Lashley (1951), which concluded that the

structure of such activities as piano playing, and in particular the verbal articulations of grammatical sentences, required 'hierarchically organised internal mechanisms' that prestructured the behaviour at various levels of abstraction.

The positions of Skinner and Chomsky at this point bring us to another impasse. Although unable to offer any demonstrations which might support his theory of Universal Grammar, Chomsky did not reject Skinner's learning approaches entirely, but neither saw any candidate learning mechanisms emerging from the behaviourist position which might reveal its aetiology. Indeed, in his own words:

> 'There are certain processes that one thinks of in connection with learning: association, induction, conditioning, ... and so on. It is not clear that these processes play a significant role in the acquisition of language.'
>
> (Chomsky, 1980: p.135)

Once again, we find ourselves in a situation for which there is no clear bridging between the end-state descriptions of 'higher' cognitive behaviours and their continuity with the more lower bound competences from which they might reasonably be expected to have grown. Further, the central part of what we might call 'learning' is more likely to be understood as the growth of adaptive cognitive structures along an internally directed course under the triggering and partially shaping effects of the environment. We have already rejected the purely associationist learning approach in the last chapter as being non-extendible, and even the more recent attempts via connectionist theories (based upon frequency of associations) have yet to offer much of explanatory value as to the emergence of a higher level of cognition. Furthermore, if language and representational processes are to be offered as the causal factors in determining the intellectual differences apparent between human and other species, are we therefore left with a need to invoke a magical explanation for its appearance ? Indeed, as Fodor and Pylyshyn (1988) were to put it:

> 'Infra verbal cognitive architecture mustn't be so represented as to make the eventual acquisition of language in phylogeny and in ontogeny require a miracle.'
>
> (Fodor and Pylyshyn, 1988: p. 40)

Learning versus representation: 'an unbridgeable gap'?

McGonigle and Chalmers (1977, 1996, 1997b) have recognised the existence of this 'unbridgeable gap in the study of cognitive functioning', in which they find themselves:

> '..... between a rock and a hard place. The rock was the bottom-up approach which has traditionally favoured the paradigms and procedures of 'the learning theorist'; the hard place was the abstract, symbol-level ridden currency of the traditional cognitive approaches which see the end product of human achievement as a domain-free manipulation of rule-based representations.'
>
> (McGonigle and Chalmers, 1997b, p. 2)

McGonigle and Chalmers go on to stress the shortcomings of the proposed conventional learning mechanisms and the more recent network approaches (including PDP), suggesting that no convincing evidence has to date been provided as to how *any* system (biological or artificial) might develop intelligent behaviour for itself, going beyond the constraints of its antecedent programmed repertoire of behaviours. For the animal learning literature, they acknowledge the contribution of Pavlov's work concerning the modified reflex, but highlight the fact that such learning processes do not appear to be 'extendible' to include the representational states of the human adult (see also McGonigle and Chalmers, 1996).

With a view to resolving this extendibility argument, there are three principle positions that one might hold. Firstly, that such an extension be an unnecessary step due to the existence of findings which suggest that much of human intelligent behaviour seems to occur without the need of high levels of cognition. Secondly, that a representational stance be supportable by models of language-like symbol manipulation without the need for associative learning mechanisms. And thirdly, a position that seeks the bridge extending across the gap between bottom-up behavioural modification and top-down representational change.

1. No extendibility necessary?

Not requiring any extendibility at all, a position may be held for which evidence for 'higher' mental processes might be deemed supportable without major revisions to associationist theory and stance. Supporters of this view can be seen to argue for the demonstration of intelligent behaviour in the absence of cognition.

In their recent book *Intelligent Behaviour in Animals and Robots*, McFarland and Bösser (1993) offer an approach to intelligent behaviour derived partly from current thinking in the biological sciences and partly from their declared distaste for anthropomorphism. They set out to avoid the temptation to interpret the behaviour of non-human animals in terms of their own subjective experience, and, when highlighting particular intelligent behaviours in a particular species, McFarland and Bösser are not saying that other species are necessarily stupid in comparison. To use their example, when comparing pigeons, moths and bats, one easily sees that pigeons are better adapted for distance navigation, whereas bats are better at aerial foraging by night, (and that moths are better at detecting and avoiding being eaten by bats). For McFarland and Bösser, it clearly makes no sense to compare the general intelligence of these animals, but it does make sense to ask to what extent their special abilities depend upon intelligent behaviour rather than upon their morphology. Their view is one of natural selection as designing agent, producing optimal adaptations in accordance with the prevailing circumstances over a sufficiently long period of time. This was not to suggest any support be given to the case for the continuity of mentation between man and the other animals, however. Indeed, McFarland and Bösser go on to explicitly list some problems for the provision of evidence for such a continuity, four of which are summarised below:

(i) that animals may exhibit considerable complexity in some respects but not others, and that different species may reach different degrees of complexity along different evolutionary routes (including convergence).

(ii) that differential brain size was not a reliable indicator of intelligence. (Coincident with the views of both Hull and Skinner above, they did not accept the evidence for intelligence to be correlated with any consistent brain allometry)

(iii) that the accounts of certain human experimental results in terms of mental representations was not well founded (e.g., there are multiple interpretations of the 'mental rotation' data). It was thus not possible to compare the intelligence of different species using standard psychological tests.

(iv) that the linking of intelligent behaviour and cognitive ability was anthropomorphic and naive. (Just as the sensory world of bats is very different to our own, so might be their mental experiences and cognitive abilities, (see Nagel, 1974)). They argued that seemingly intelligent behaviour could occur in the absence of cognition.

Intelligence without cognition ?

McFarland and Bösser also realised that although the finer details of behavioural control mechanisms might change with our gaining new knowledge of their function(s), the adaptive significance of behavioural change was likely to remain more stable. In their subsequent attempts to outline a working definition of intelligence, McFarland and Bösser (1983) follow Tuddenham (1963) who came to the conclusion that:

> '... intelligence is not an entity, nor even a dimension of a person, but rather an evaluation of a behavioural sequence (or the average of such), from the point of view of its adaptive adequacy. What constitutes intelligence depends upon what the situation demands.'
>
> (Tuddenham, 1963: p. 517)

This view was endorsed by Hodos (1982), who promoted the idea of animal intelligence as providing an abstract characterisation of an individual's behavioural responses to pressures received from the environment. A final comment on the contribution of McFarland and Bösser at this point, concerns the issue of using human intelligence as a basis for understanding the intelligence of other species of animal: as is true of animal behaviour, many aspects of human behaviour do not involve cognition. To take another of their examples again, the system that controls spatial orientation in humans is a highly sophisticated, pre-wired adaptive control system (Howard, 1982), and the behaviour it controls would seem to be highly intelligent if performed by a robot. In conclusion, McFarland and Bösser (1993) thus offer reasons for abandoning the quest for a general animal intelligence, and argue instead for a shift of focus

from the search for a comparative intelligence *per se*, towards the study of intelligent behaviour.

2. Representational stances with a disdain for reflexive associationism.

In their review of representational stances in cognitive psychology, McGonigle and Chalmers (1997) discuss the appearance of 'a "new" system competence based on the internal manipulation of language-like symbols'. For such a system:

> 'Its properties are that it provides for a connectivity through meaning alone, affording the possibility of 'pure thought'; the conjuring up of possible worlds; judgement and reasoning through propositionally driven hypothetico-deductive mechanisms; the creation of a mental world 'abstracted' from the space time reality in which we have our existence.'
>
> (McGonigle and Chalmers, 1997: p. 5)

If these symbol-user properties are those which are to partly explain the intellectual differences between non-human species and linguistic humans (as McPhail (1982) and others suggest) then a rather large onus is placed upon those strongly arguing for a representational account of cognitive development in support of their position. Often ignoring issues of biological (evolutionary) development, a fundamental difference in gathering their evidence is that in contrast to the purely behaviourist tradition (according to which, both stimuli and responses be directly observed and measured), the study of representational concepts calls for a method of indirect evaluation, requiring the use of inferential processes. For Karmiloff-Smith (1992), for example, the study of development as a theoretical tool for exploring the human mind from a cognitive science point of view is a methodology. Taking a developmental perspective of human cognition, her objective is to better understanding the 'built in' architecture of the human mind, the constraints upon learning, and how knowledge changes progressively over time, in the hope of finding subtle clues to its final representational format in the adult mind. Karmiloff-Smith's results lead her to the conclusion that children's learning of language, numeracy and drawing skills (in early skill acquisition) follow a common developmental pattern explained by a process that she has

termed 'representational redescription' (Karmiloff-Smith, 1991, 1992, 1993).

Representational redescriptions

Having asked children of different ages (four through eleven years) to draw a picture of a house, for example, and then of 'a house that doesn't exist', she found that for the latter task, the younger subjects performed differently from their elders. Whereas the four to six year olds typically changed both relative sizes and shapes of their drawn houses and left out familiar elements from the end of their usual drawing sequence, they did not reposition any elements or delete elements normally occurring in the middle of their drawing sequences. The theory of representational redescription proposes that this result should be explained in terms of the spontaneous and endogenous construction of increasingly explicit internal representations from what was initially a unified procedure. The growing ability to alter the distinctive elements of familiar pictures and to change their sequence of occurance whilst drawing them, is said to be correlated with their progressive "availability as data" as parts of the procedure. This then makes them amenable to comparison and alteration. Karmiloff-Smith (1993) goes on to conclude that there are at least four levels of representation in the human mind (procedural, linguistic and two unconscious levels) and that:

> 'The model of representational redescription postulates that the mind stores multiple redescriptions of knowledge at different levels and in different types of representational format, which are increasingly explicit and accessible. At the initial level, I argue that representations are in the form of procedures for responding to and analysing stimuli in the external environment..... The redescriptions are abstractions in a higher level language and are open to potential intra- and inter-domain representational links, a process which enriches the system from within.'
>
> (Karmiloff-Smith, 1993: pp. 597-598)

However, as discussed above, she at no time specifies the primitives (initial descriptors or reflectors of world events) required by this theory, and as pointed out by Hendriks-Jansen (1996), locates her explanation of development entirely 'within the subject's head':

> 'The process of representational redescription is assumed to be endogenously driven, and the original procedures, as well as the various levels of representation that result, are all assumed to be internal to the creature.'
>
> (Hendriks-Jansen, 1995: p305)

As with so much of the ape-language work above, and indeed many of the cognitive models of human language acquisition in psycholinguistics (see Garnham, 1986, for review), most theorists have presupposed the existence of some kind of rule-set(s) that generates the representations required of the putative black-box of structures inherent in the detail of their particular theories. If, from a representational stance, our only recourse to discovering the details of the process is to derive it from its expression via the language of subsequent, and end-state redescriptions, we are brought no closer to an understanding of their development from the viewpoint of representational redescriptions. Such inferences concerning the structure and function of representations remain based upon an analysis of the effects that representations appear to have on subsequent behaviours. As a result, therefore, because cognitive development involves learning and thinking processes that are not directly observable, it will be necessary (even for human subjects) to design and implement experiments employing known, identifiable stimuli, that elicits a response in the subject which demonstrates the use of a previously perceived and stored representation. Only by so doing might one be able to sufficiently ground a representational theory in order to incorporate it into a more empirically founded explanation of cognitive growth and development.

3. A bridge between behavioural learning and representational stances.

In an attempt to forge the gap between the associationist and more representationist schools of thought concerning the development of 'higher' mental processes, there are a number of researchers who have sought to explain its origin not in terms of learning mechanisms *per se*, but in biological (evolutionary) development. From this perspective, it should be as true of cognitive structures that they have evolutionary antecedents, as would it be for the modern mammalian eye. This viewpoint requests the transfer of questions of development away from its purely

psychological abstractions to a more firm grounding in biology. In the words of Noam Chomsky:

> 'There is.... no reason to adopt the common view that the human mind is unique among the systems known to us in the biological world in that, in its higher cognitive functions, it is unstructured apart from some minimal "hereditary forms" or "quality space".'
>
> (Chomsky, 1980: p.213)

In his *Rules and Representations*, Chomsky (1980) supported the view (as did Lenneberg (1967) before him) that the origin of the 'human language capacity' be treated in the same way as any other physiological organ of the body. Further, if language was to constitute an innate process (as was the case for Chomsky's concept of Universal Grammar), it was pertinent to ask the question as to what exactly it was that was postulated to be innate for linguistic behaviour ? Chomsky accepts that:

> '[As] the rules of grammar enter into the processing mechanisms, then evidence concerning production, recognition, recall and language use in general can be expected (in principle) to have bearing on the investigation of rules of grammar, on what is sometimes called "grammatical competence" or "knowledge of language".'
>
> (Chomsky, 1980: p.200)

He then went further to comment upon investigation in psychology 'which in contrast, is more concerned with performance, not competence'. As we saw above with the example of Karmiloff-Smith and her position's inherent difficulties with theoretical grounding, much of psychology is concerned with processes of production, interpretation and suchlike, 'which make use of the knowledge attained, and the processes by which transition takes place from the initial to the final state, that is, language acquisition.' What Chomsky does not tell us, however, is how one might go about discovering the 'innate mechanisms' he believes might be supporting his notion of Universal Grammar. Neither does he speculate as to how one could embark upon a program of research which might identify the biological principles that underlie language use. At best, we are presented with a vote for a method without a procedure, a desire for an investigation of the origin of language use for which:

> 'The abstract study of competence systems and the study of process models offers a great deal of promise..... and can place significant conditions on the biological mechanisms that enter into the language capacities.'
>
> (Chomsky, 1980: p.216)

Again, Chomsky does not go on to offer any indication as to how one might proceed. He does, however, envisage barriers to progress for the elucidation of the biological basis of human language processes, including the 'impossibility of direct experimentation', which might otherwise have helped answer some of the many questions he raises.

A new genetic epistemology

One further line of research which also supported an approach allied to the biological origins of cognitive growth and development, was that put forward by Jean Piaget in the latter part of this century. Although Piaget was to have many a battle with Chomsky concerning the details of the nativism of Generative Grammar, he shared with him the desire to ground the questions of higher mental processes in evolutionary biology. Centred principally around the development of intelligence in human children, the work of Piaget (1971) is also noteworthy in that it might at first appear to offer a reasonable model for addressing issues in animal cognition and the differences between human and non-human species. With regard to cognitive development in humans, Piagetian theory rests on the:

> '..... mutual interrelation of schemes and on the differentiation and enrichment they undergo by being constantly adjusted to the external world.'
>
> (Etienne, 1973: p.376)

By the word 'schemes' here were meant the underlying structures and organisers of actions which are involved in the adaptation of an individual, to the environment in which it is situated, and through the interplay of what Piaget calls 'assimilatory' and 'accommodatory' processes. According to this view, adaptation was not merely the emergent unfolding of totally preprogrammed structures, but neither was it a response absolutely determined by environmental factors. The adaptive response was said to be a consequence of both internal factors (the 'schemes' and their structural organisation in the brain) and external

pressures (constraints imposed by the environment). Continuing with Etienne for an example:

> '... a baby who sucks the nipple of a bottle or the corner of a blanket incorporates or *assimilates* these external elements into his sucking-theme; in other words, he applies a pre-existing general action-pattern to these objects.'
>
> (Italics preserved, Etienne, 1973: p.375)

The first action schemes here (sucking, prehension, visual fixation etc..) have their origins in the reflexes present at birth, and it was Piaget's contention that cognitive growth and development consisted in the continuous differentiation, generalisation, and co-ordination of these reflexes into organised actions (or sequences of actions) composed of those same 'primitive' reflexes. Throughout human cognitive development to adulthood, these processes were believed to continue extending from these relatively simple sensory-motor accomplishments towards the higher symbolic manipulation systems of complex mathematics (Piaget, 1971).

Stages and benchmarks of cognitive growth

The results of Piaget's investigation of the human child's intellectual development, lead him to distinguish four main periods which, although varying in age of onset for the individual child, would be sequentially dependent in their order of appearance. These were (see Piaget, 1950; Piaget and Inhelder, 1969): the sensorimotor period (from birth to two years), the pre-operational period (from two to six or seven years), a period of concrete operations (six to eleven or twelve years), and finally a period of formal operations (from the age of eleven to twelve years of age onwards to adulthood).

The sensorimotor period

According to the developmental theory of Piaget (see also Flavell (1963) for review), the sensorimotor period comprised six main stages. At first, reflexive 'schemes' were said to be applied to a variety of objects in the environment, primarily through the actions of sucking and other oral contact stimulation. These reflexive behaviours then led to the formation

of self-stimulation habits that concerned the infant's own body. This second stage would include such behaviours as repeated thumb-sucking and visual tracking of the moving limbs. At this stage, Piaget and Inhelder (1969) define what they call the stage of 'the primary circular reactions', in which a circular reaction involved the repetition of a particular behaviour that had a pleasant effect for the infant. With an onset at about four to eight months of age, the child was said to begin exhibiting behaviours more oriented towards objects situated further afield in the environment. These were called 'secondary circular reactions', and accompanied the production of new behaviours made possible by the co-ordination of vision with prehension. For example, a six-month old child would pull on its blanket whilst watching it move, or repeatedly shake a rattle whilst attending to its intermittent sound. Stage four was typically reached as the child ended its first year of life and was characterised by the application of complex co-ordinated sequences of actions directed toward objects in the environment. This process saw the beginnings of the so-called 'tertiary circular reactions' which now appeared to involve the more systematic and experimental trial-and-error manipulation of objects. At this stage the child would typically be seen to watch the effects of repeatedly dropping objects or the repetitive placing of objects in and out of containers. These different 'circular reaction' stages were said to come together with each other in stage five, paving the way for the progressive mastery of the spatial and causal relations between objects. Finally, for stage six, the earlier stages were said to be combined again, but this time 'mentally', so that a given 'scheme' (for example, the removal of an obstacle) could be used to in some way 'activate' another scheme (say, the discovery of a hidden object). Cognitive development during the first two years of life consisted for Piaget, of the progressive establishment of successive invariants. One of the main invariants said to be acquired during this sensorimotor period was the phenomenon of 'object permanence', a term indicating that the child was thought to conceive of objects as fixed and permanent entities.

Object permanence

This ability to understand that an object existed even when hidden from view, was assessed by examining an infant's reaction to objects that were moved about and hidden whilst observing both the movement and the

hiding actions of the experimenters (Piaget, 1942, 1954). Marked as one of the most important achievements for the growth of intelligence, the development and construction of object permanence is of great relevance because 'the object' always played the role of invariant in all contexts of the infant's cognitive acquisitions (for example, the organisation of space, time and causality). For Piaget then, by the end of the sensorimotor period, the permanent object became the first invariant of "the practical group of displacements" whose role was to organise the child's movements in space and to structure the movement of objects in the external environment.

Post-sensorimotor periods

Following Piaget's sensorimotor period, the infant was said to enter a pre-operational period extending from two to six years. This later period was characterised by the appearance of new types of behaviours, including a semiotic, or symbolic function (Piaget and Inhelder, 1969). This new function described the ability of the two year old child to represent an object or event (a "referent") by novel means (called "signifiers"). This signifier might be a gesture, word or mental image made possible by the use of symbolic tools (language, representative and conceptual thinking), but its importance as a defining benchmark of intellectual growth, was that these achievements were those thought to be the precursors to the realisation of simple inferences and figural categorisations (Bruner et. al., 1996).

By the time of reaching the age of six to seven years, the child was said to enter a period of concrete operations, a stage in which number concepts and the principles of physical invariance (e.g., weight) were mastered (Piaget and Inhelder, 1969. The benchmark here, was that the cognitive abilities now extant contained structures which supported more complex operations such as those based upon class inclusion, mental reversibility and transitivity. The child's application of this logical framework was, however, said to be limited to the outcome of actions upon objects, in the sense that intellectual operations were still strongly linked to objects and, importantly, relations among objects (Inhelder and Piaget, 1964).

The final period of Piaget's theoretical sequence of stages in the human child's intellectual development was the period of formal operations,

lasting from the age of about eleven years until adulthood. It was in this latter period, according to Piaget, that adolescents adopted logical strategies that were to some extent independent of their content. Thinking was said to occur on a hypothetico-deductive level wherein mental operations combined together and were applied to a content that had a propositional and a hypothetical status (Inhelder and Piaget, 1964). Having attained this final level of competence, the child was thus capable of drawing "if-then" conclusions, the basis of the ability to assess the status of possible events and outcomes 'in the mind's eye'.

Problems of Piaget's genetic epistemology

At first glance, Piaget's theory would appear to provide a framework which suggests how one might address the problem of the extendibility of basic learning mechanisms to the more complex domains of complex problem solving in 'higher' cognitive processes, but it still offers no detailed explanation concerning where they 'come from' (see Fodor and Pylyshyn, 1988; Newell, 1992) in contrast to the earlier, more associationist models of behaviour. Having remained an ongoing problem for researchers coming from quite different perspectives (Karmiloff-Smith, 1992; Hendriks-Jansen, 1995), as McGonigle and Chalmers were recently to point out:

> 'Piaget's epigenetic approach has failed to identify - even admit there could be - 'structures unpolluted by history and genesis', i.e., the lower bound primitives of the system, and has left this approach with an impossible task of having to specify the transition within ontogeny from 'habit of mind', arbitrary to non-arbitrary and necessary knowledge.'
>
> (McGonigle and Chalmers, 1997: p.4)

However, Piaget's more formal axiomatic-based top-down ontology does nonetheless offer a model for which there is an identifiable empirical program for establishing the origins of 'necessary knowledge' (Smith, 1993). But, whereas on the one hand Piaget views knowledge as a product of biological adaptation that is constructed as a result of each individual's continuing interaction with the environment, on the other hand, his model describes the development of mental mechanisms ('schemes') and mental operations that have their origins in the (unspecified) innate

reflexes of the newborn. As McGonigle (1996) writes in his review of Smith (1993):

> 'To avoid nativism, or knowledge as recollection in the Platonic sense ..... Piaget emerges diminished with a "bad explanation in response to the problem", but better than nativist alternatives, which are not deemed explanations at all'
>
> (McGonigle, 1996: p.2)

Indeed, even Chomsky (himself accused by Piaget of postulating a "biologically inexplicable" theory in Universal Grammar) criticises the Piagetian school for not supporting their original claims for a biological origin for cognitive development:

> '.. no matter how substantial the evidence ..... and no matter how weak the argument for ontogenetic development, nevertheless we must maintain the thesis that the principles in question are derived from "regulatory or autoregulatory mechanisms" in accordance with the hypothesis of "developmental constructivism".'
>
> (Chomsky, 1980: p209)

What Piaget had actually done, was to divorce his logical explanation of the processes developing in the 'epistemic agent' through accommodation and assimilation, from the biological epigenetic account of the possible structures involved in the growth and development of cognition. Indeed, although Piaget was addressing the need to study the development of life-historical competences, on an individual subject basis, and in the context of a more dynamic-interactive theoretical framework, his experimental procedures were never capable of revealing what the structures motivating cognitive growth might be. But although he remained unable to shed any light on this issue, Piaget had nonetheless contributed much of significance in the course of his studies. He had shown that during the early stages of growth in human cognitive performance there were a range of identifiable competences which an individual subject might be capable of, and he described their coming into being along an invariant sequential pathway throughout intellectual development from infancy to adolescence.

Given these considerations, although now motivated by a more ontogenetic stance, placing emphasis once more upon the life-historical development of the individual subject, it still remained to be seen what particular kinds

of experiments might be designed and implemented in an attempt to reveal such evolutionary processes as might have been involved in the development of complex cognitive functioning. With a view to addressing these concerns, and in an attempt to bridge the gap between studies of bottom-up tightly-coupled reflexive animal behaviours and the more top-down approach often employed in explaining human 'higher cognitive' functions, it is now time to turn to the work of McGonigle and his colleagues who, for over a quarter of a century has been exploring these issues his studies of comparative learning and development.

Chapter 3

A new agenda for comparative learning and cognition

For much of the literature concerned with the development of human cognition, McGonigle and Chalmers have continually emphasised the recurrent problems inherent in the use of experimental paradigms which heavily invest in complex, language-like abstractions (wittingly or unwittingly). Such procedures have in common the fact that they may only be operated upon by an experimental subject which is already symbolically competent. The important consequence of these factors for any comparative animal cognition study, is that the currency of explanation for many of these cognitive behaviours will start from a point beyond which most learning accounts finish, that is, the currency of symbol manipulation (a recurrent problem also for the interpretations of the ape-language and 'theory of mind' work discussed in the first chapter above). If this is indeed the case, how then might one be expected to evaluate the existence of any 'higher cognitive' functions in the non-human animal ? The work of Piaget, although significant, has been of no help in this regard. His fame rested upon his developmental psychology, and in raising the status of the idea that development of any kind, not just psychological or behavioural, might be better understood in terms of ontogenetic (i.e., individual life-historic) evolutionary processes. Although Piaget's genetic epistemology centred around structuralist ideas of organisation, self-regulation, co-ordination and construction rather than the conventional evolutionary concepts of variation and selection, he was nonetheless unable to produce any convincing demonstrations of symbol-based cognitive development which did not presume a linguistic agent. By remaining insensitive to those subjects lacking in 'semiotic instruments' (Piaget, 1971), such evaluations would necessarily imply in-principle limitations upon the performance of both young children and non-human species alike. Also, although the studies of the Genevan school have been illuminating, the constructionist model itself is difficult to assess, because it remains at the level of metaphor. Further still, the interpretations of Piaget, which suggest that a child progresses through a regular sequence of cognitive stages, do not include any mechanism or principle which might explain why the child moves from any given stage to the next rather than to some quite different stage.

A paradigm is therefore needed within which complex cognitive skills may be fractionated into its component parts in such a way as to allow their subsequent recombination in tasks sensitive to the dynamic-interactive cognitive growth of the individual subject. What is required here, is the development of problem-solving tasks which are sensitive enough to inform both subject and experimenter of any developmental learning changes taking place, whilst at the same time do not presume the possession of representational or symbolic devices, the detailed evolution of which themselves remain to be explained. Motivated in part by Piaget's genetic epistemology, the solution explicitly argued by McGonigle and Chalmers (1977a & b) satisfies four principle requirements for the establishment of a common currency of tasks and measures for comparative and developmental approaches to cognition. These were: (i) the provision of a specification of the design primitives to be used in order to enable the plotting of any growth trajectory evident for a given competence, (ii) an evaluation of the emergence of complex cognitive functioning from these design primitives derived from paradigms which monitor cognitive growth with a high density of sampling and measures of complexity, (iii) subjects allowed the opportunity to self-regulate their behaviour over long periods of time in response to increasing task complexity, (iv) the use of non-linguistic tasks neither requiring, nor presupposing, any degree of linguistic competence on the part of the subject for their solution. In order to have addressed both the theoretical and methodological criticisms outlined above, whilst simultaneously assuring the most optimal animal laboratory conditions and procedures as reviewed earlier in the chapter, some of the necessary preconditions for the production of candidate behaviours for the study of a new era in comparative animal cognition have now begun to be revealed.

Rationality in human and non-human primates

These four points above all propose seemingly rational ideas, and indeed, are indicative of rational thinking in any test subject at first glance; but they are very difficult to formalise. Rational thought may be expressed in terms of (the) rules of logic, probability or statistics, signal detection theory and more recently, complexity and chaos theories, but all of these approaches remain to some extent incomplete in their detailed application.

One might also propose that rational thought be based upon deductive and inductive reasoning, but even here there are no firm promises (as one is so often made painfully aware); rational thought does not necessarily guarantee rational behaviour. However, if rational action is to be determined by rational thought, then one might expect to find that individuals incapable of rational thought would also be incapable of rational action. Suppose, for example, that a subject is told that A > B and that B > C, then asked A ? C. One would expect a normal human adult to realise that indeed A > C - solving a 'transitive inference' problem. If repeatedly failing or showing inconsistency with such a problem, one might say the subject was not behaving rationally. But how might one address this same question to a non-linguistic animal ?

A new cognitive learning approach

The particular scenario proposed and investigated by McGonigle and Chalmers (1996, 1997a, b; McGonigle, 1991) is that of a behaviour-based cognitive learning approach which focuses upon 'spontaneous change derived from cognitive self-regulation and self-organisation' in the laboratory subject. Motivated by 'Piaget's constructivist, epigenetic conception of rational systems from an 'adaptive' perspective', the experimental approach proposed by McGonigle and Chalmers (1997b) aims to produce an empirical 'characterisation of cognitive complexity' by the use of three main cognitive tasks: seriation, transitivity and classification, all reflecting central features of adaptive cognitive architectures. In agreement with Anderson (1990), McGonigle and Chalmers 'subscribe to the notion that an evolutionary optimising principle is to get the most behaviour for the least effort' and that therefore:

> '... assumptions about computational limitation provide an important rationale for cognitive growth and enable the agent to utilise a new form of (internal) arbitration based on criteria of cognitive economy which favour procedures which achieve success with the least investment in cognitive resource. And it is from arbitration of this sort that we believe cognitive autonomy and crucial aspects of self regulation in advanced primate systems evolve.'
>
> (McGonigle and Chalmers, 1977a: p1)

In their ongoing program of research, therefore, McGonigle and Chalmers seek to assess behaviours 'that are not behaviour trouvé... but rather behaviours which are constructed over the task history and... life history of the agent within scenarios tailored to examine the dynamics of cognitive regulation'. Such an experimental agenda sets itself the aim of seeking evidence for an agent's use of cognitive economy as a data reduction strategy. Using design primitives which are characteristically different from those classically employed and explained by sensorimotor or associationistic mechanisms, McGonigle and Chalmers (1997a) report upon a series of experiments conducted with both human and non-human primates which have provided for a rich dynamic exchange of activity between task and subject (so provoking adaptive responses in the face of either success or failure), using tasks which can in principle be extended to infinite levels of difficulty by the continuing use of increasingly larger search spaces.

Working simultaneously with human children and laboratory monkeys, McGonigle and Chalmers (1977a) were the first to explore the possibility that a non-human primate (in this case, a squirrel monkey) might be capable of transitive inference. Using a variation of a non-verbal task (involving coloured tins of different weights instead of wooden rods of differing length) originally designed for use with children (five-term series, Bryant and Trabasso, 1971), they reported that the monkey's performance on test trials were closely comparable with those of the human four-year old children (which, according to Bryant [1974, p.47] '.....demonstrates conclusively that young children are capable of making genuine transitive inferences'). Later, using both verbal and non-verbal versions of the same task, McGonigle and Chalmers (1980, 1984, 1986) were to report that on all major points of comparison, the monkeys were almost identical in profiles when compared with the performance of young human children. Further, taking the finding of a symbolic distance effect as evidence for a 'spatial paralogical device' in human adults and older children, the monkeys (subjects unable to perform formal logical tasks) were seen to perform comparably well. Following these results, McGonigle and Chalmers (1986) came to the reasonable conclusion that the ability to order items transitively was a 'pre-logical phenomenon'. However, knowing that young children, squirrel monkeys and even pigeons (Terrace 1987, Terrace & McGonigle, 1994) can, with training, solve

transitive inference problems, they evidently do not do so via syllogistic reasoning. So, although continuing to remain sceptical of the claims for rational thought as being a necessary precursor to rational action, it can be seen that ordering skills may be assessed transitively according to carefully designed experiments with non-human, non-linguistic subjects.

In the light of this finding, what then of the claims for the young child's performance ? Were they not so capable of such abstract thought as previously claimed (Bryant, 1974), or indeed, could it now be said that the monkey all along be a doyen of logic despite the lack of any 'semiotic instruments' or cogent linguistic expression ? In order to address this question, a series of experiments were begun following the relational primitive work of McGonigle and Jones (1978). The problem still in search of a solution, however, was that of the choice design primitives to be used in any attempt to provide an explanation of their incorporation into an epigenetic growth trajectory which might result in the development of the evolving cognitive structures seen in the human child maturing to adulthood.

Relational competence: a qualitative shift in cognitive growth

As discussed in reviewing Harlow (1949) in the previous chapter, if animals can remember at the start of a trial the outcome of the previous trial, and if they also identified stimuli on the basis of whether or not they had been previously selected, then they should be able thus to master learning sets. Indeed, the success of this strategy depends upon the animals remembering a good deal of information from one trial to the next. However, exactly what it was that was being learned in his learning set experiments remained poorly characterised by Harlow, and we inherit a largely descriptive account of his nonetheless intriguing results. Unfortunately, the paradigm that Harlow employed underwent no further development and the degree to which the learning sets might be extendible were not explored. New stimuli were used over hundreds of trials, but choice responses continued to require binary decision matrices for the solving of one-trial learning, reversal or delayed matching-to-sample experiments. An inductive mode of reasoning could always be offered to explain most successes - generalisations being formed through the perception of regularities in past trials being used to predict similar

regularities in future trials - but this was true only in so far as generalisations could be used as predictors varied according to perceived causal links. However, little evidence was provided for *reasoning* in the learning set experiments. For Harlow's non-human animal work, associations require contingency (not simple contiguity) between events and objects, reflecting a genuine causal relationship in principal, but this did not require the possession of a concept of causality *per se*. Further, as with the work of Köhler (1925), without closely controlling for a subject's potentially relevant previous experience, any analysis of problem-solving was weakened. One thus remains at this stage without any coherent view of reasoning, let alone rational thought, in the non-human animal. What was still required was the design of simple problem-solving experiments which might deny "classical" sensorimotor-reflexive interpretations of behaviour, yet at the same time appeal to a non-linguistic explanation of the human comparative data.

Binary discrimination learning

In experimental studies of binary discrimination learning, the occurrence of systematic response tendencies have been claimed to be indexical of an 'organised lawful process' taking place prior to the acquisition of success (Krechevsky, 1932). Further, whilst offering a continuity interpretation of his findings, Spence (1949) insists that:

> '.... so long as the subject is receiving discriminably different stimulation from the positive and negative discriminanda during the presolution period, differential associative tendencies will be developed with respect to them.'
>
> Spence (1949: p. 719)

According to this view, Spence was rightly arguing for an incrementing excitatory (physiological) strength coming to be attached to an absolute stimulus value during the rewarding of successful discrimination learning, as compared to the competing (non-rewarded) negative value. Such a contention is crucial to our understanding of an animal's generalisation of trained discriminations to novel stimuli sets in terms of a model of self-regulated differential reinforcement. The difficulty here for the subject, is to somehow become attuned to the salient characteristics of the stimulus array in order that differential reinforcers be discriminable.

For example, much of the experimental work conducted by Spence (1945) employed the Lashley type of jumping apparatus (see figure 3.1), in which the subject learns to jump across a space to one of two windows (W), the approach to one but not the other leading to a reward. Acquiring the habit of fixating the lower ledges to which it must jump (L), the fixations bring discriminable stimulus cards into the visual field of the subject (although there is no guarantee that discriminably different retinal patterns are received from the stimulus array). With sufficient training, animals will approach the correct stimulus on the majority of trials. Krechevsky (1932) and Lashley (1929) had proposed that such successful performance depended upon the subjects' 'hypothesis testing', for example, "go left", "approach the black one", and so forth. This, they contend, may be done without their acquiring any differential response tendency to the absolute stimulus value until the time of solution, appealing to a non-intertrial theoretic interpretation. In contrast, Spence (1945), (see also McCulloch and Pratt, 1934) argues for a continuity theory interpretation of discrimination learning as discussed above.

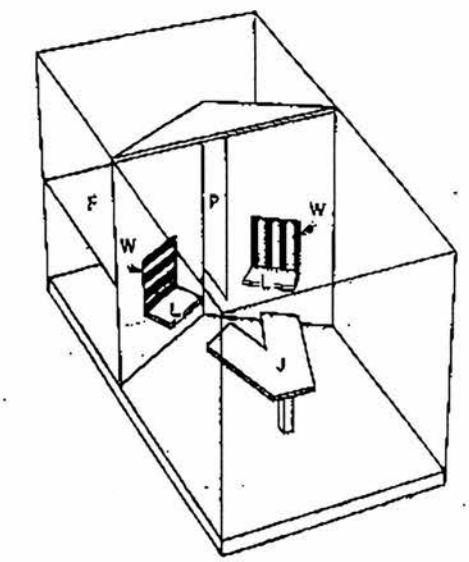


Figure 3.1. A typical jumping stand used for studies of discrimination learning. (after a drawing by Sutherland, 1964)

If Spence was correct, the jumping-stand experiments are to be explained thus: when an animal was responding with a "going-left" hypothesis, it received reinforcement every time the positive cue (e.g., horizontal striped card) was on the left side, but was not reinforced when the negative cue (e.g., vertically striped card) was on the left. After a number of trials with such differential reinforcement, the horizontally-striped

cue acquired a greater exitatory strength than did the vertically-striped one, but initially the subject continued to choose the left because the difference between the left and right *positional* cues was greater than the difference in excitatory strengths of the discriminable absolute stimuli cards. Only later, after continuous reinforcement, did this latter difference become more salient over the positional cues; the subject then abandoning the 'left-going' hypothesis for the more reliable (and indeed, 'correct') 'horizontal striped card-going' hypothesis. The strengths of these two opposing theories may be assessed by looking at the ways in which an animal would deal with situations involving continuous serial-reversal (see fig 1.7 (b)). According to a more associative, non-continuity theory, initial training should not effect the subsequent learning of the reversal problem. The implication for the continuity theory was that the learning of such a reversal group would be slowed in comparison with a control group receiving merely a fifty percent reinforcement schedule. Strong evidence has been provided (both at the time and subsequently) to support the continuity theory and it has repeatedly been reported in a number of studies (Spence, 1945; Krechevsky, 1938; McCulloch and Pratt, 1934; Schrier, 1966). However, these results revealed something more about the problem solving abilities of animals than could ever have been shown by more conventional discrimination training. This performance could not be explained in terms of association formation, wherein the stimuli concerned are those confronting the animal, even when attentional processes were taken into consideration. In order to account for successful serial reversal learning in terms of association formation, it must be assumed that the memory of the events of one trial serves as a cue for the response on the subsequent trial. Such a position offers support for the claims made for the formation of learning sets in the preceding section, i.e., that an animal remembers a good deal of information from one trial to the next.

Absolute versus relational object discriminations

One prediction of this model is for the ability of an animal to generalise a previously trained discrimination by transferring its logical operation to new sets of previously unseen stimuli for the purposes of further discrimination. Such two-stimulus transpositions have been recorded with every species so far tested (including honeybee, wasp, fish, various birds,

rodents and primates) but intermediate-stimulus transposition appears to be rare (Reese, 1968). This latter observation brings us to the final point in this section, which draws out the implications of an important distinction to be made between absolute and relative discrimination problems. The terms 'absolute' and 'relative' discrimination problems require three stimuli, *A*, *B*, and *C*, ordered according to some determinable diminution. An intermediate stimulus *B* is paired with *A* on fifty percent of the trials, and with *C* on the remainder. For the absolute discrimination problem, subjects must choose *B* whether it be paired with either *A* or *C*. For relative discrimination problems the subject is required to choose *B* when *B* is paired with *A*, and to choose *C* when *B* is paired with *C*. In reviewing the results of numerous experiments which had investigated both absolute and relative discrimination problems using the same stimuli (usually involving brightness relata), Reese (1968) reports a consistent difference to be found. Performance on the relative discrimination problems were typically higher than they were for trials involving an absolute discrimination problem (for example, 96% Vs 51% correct responses). Summarising the evidence for the relative ease of learning relative and absolute discriminations, Reese (1968) states that:

> 'The absolute discrimination problem is harder to learn than the relative discrimination problem; but it can be solved, and the solution is apparently based on absolute properties and not on combinations of relative properties. Even on the relative discrimination problem, absolute properties have an effect on performance. The difference in difficulty is attributable to the occurrence of transposition, which facilitates performance on the relative discrimination problem and interferes with performance on the absolute discrimination problem.'
>
> Reese (1968: p. 198)

Despite this consistent finding however, the relative discriminations are harder to learn than are the standard simultaneous discriminations of Harlow, due to the transpositions in relative discrimination problems being corrupted or interfering with absolute responses transferring between stimuli sets. Using squirrel monkeys as subjects, McGonigle and Jones (1978) confirmed the findings of relational discriminations outperforming conservation of absolute stimulus discrimination (with both brightness and size relata conditions). The robustness was even found to hold in the face of introduced contextual variation: background illumination for the brightness, and set expansion for the size conditions

respectively. These results led McGonigle and Jones (1978) to propose that finding such relational encoding in the monkey was evidence of its being a design primitive, not reducible to any lower layer of competence accountable for by any of the traditional learning theories. Although as long ago as 1884 Romanes had proposed that relational property evaluation might lead to ideational thought, this work was to provide the first empirical demonstrations of the idea in a non-human animal.

Relational primitives

This set of findings continues to lay the foundation for the rationale of extending the learning set results to include investigation of longer string lengths and the differentiation of absolute and relational discrimination problems as proposed by McGonigle (1990). It is now quite clear how a case for the continuity of mental processes across species may be supportable by the use of consistently non-associative interpretations of learning set results, serial-reversal and relational learning discriminations, even for the non-verbal subject. Considered as primitives, McGonigle and Jones (1975, 1977, 1978) argued that perceptual relations were not merely demonstrations of abstract discriminatory competences, but provided evidence of on-line dynamically interactive learning processes in action. Indeed, as McGonigle and Chalmers (1997b) were later to argue:

> 'Only by comparing and contrasting stimuli from a known set can the defining features logically and inductively be determined.'
>
> McGonigle and Chalmers (1997b: p. 10)

In other words, relational competence makes perfectly rational adaptive sense when situated in the context of certain problem solving environs. The significant finding of McGonigle and Jones (1978) was that their monkeys' reporting of relationally encoded stimuli was 'easier' than for their encoding of the absolute stimulus properties of array discriminanda. Furthermore, based upon the trial-one results (in which only the relational encoding subjects could immediately predict the correct choice) a distinction was made evident between the characteristics of non-specific transfer, based on non-arbitrary relational rules and non-specific, learning to learn effects. Although the monkey did not approach the performance of human subjects at that time (in terms of the numbers and

'levels' of categories used), in their discussion, McGonigle and Jones suggest that:

> '..... the possibility remains that, given scope for *multidimensional* determination of a stimulus configuration, the monkey will learn to identify many individual configurative patterns as unique ones.'
>
> (Italics preserved. McGonigle and Jones, 1978, p.659)

Seriation and transitivity

The later studies of McGonigle and colleagues, were designed with a view to determining the degree to which multiple relational codes might be derived from such primitives at the perceptual (rather than at the semantic) level of functioning. As may be seen from the earlier discussions of the ape-language work above, a subject could only be said to have learned to identify the (trainer's) concept only if that concept was putatively included among those the subject already entertained. Levine (1971) had already argued this with admirable directness, and although his conclusions were based upon the use of simple features (red, square, larger, etc.,) they apply equally when names are taught for the more complex classes, relations and logical connectives. But here too, 'names' could only be associated with the target concept if the concept is included amongst those the subject entertains, a phenomenon not previously made available for empirical analysis. The move towards a transitive-choice task was motivated by the desire to elicit in the subject a manipulation 'in the mind's eye' from a choice array of items in ways not cued by stimuli position or absolute features alone. Allied to the position of the mid-twentieth-century Piagetian psychology, and following the construction of absolute object relations at the concrete level, the later cognitive stages ascribe the development of conservation skills, reversibility, and logical operationalism to a level of 'private thought' as exemplified in the classic syllogistic inference of Inhelder and Piaget (1964): Edith is fairer than Suzanne; Edith is darker than Lilli; Which is the fairest ? Here the triadic relationship needs be constructed from a given of two diadic relations with an overlapping connective. The solution is thus not present in the givens by inspection. What is required in coming to a solution is for the subject to construct a mental series in which A > B > C (or A < B < C). Using this idea as an index of relational-based object coding in the monkey, McGonigle and

Chalmers (1977) went on to determine whether there might be any evidence for principled relational understanding in the non-human primate. Using the five-term series of Bryant and Trabasso (1971) mentioned above, in which four connecting pairs are trained (A > B, B > C, C > D, D > E) followed by a test probe pair (B ? D) unbiased by either end-anchor or other 'learned' transitive outcome, the first demonstrations were soon forthcoming. But had the monkey recombined the paired items in order to form a principled series A -> E (or E ->A) ? On all points of comparison explored, the monkey and child performances were highly similar in profile, and even with the later child studies exploring the detailed effects of linguistic and non-linguistic versions of the monkey task (McGonigle and Chalmers, 1984), the results were highly congruent across both species and task type.

A further measure of the degree to which a linear representation might be constructed by the subject in these experiments is suggested by Potts (1972) following his discussion of symbolic distant effects (SDE). For both the human and monkey subjects, McGonigle and Chalmers (1984, 1992) report classical SDE functions for children as young as age six, and also, for the very first time, with the non-human primate. The interpretation now offering itself, was that the monkey had indeed integrated pairwise items into a representation of a series. These findings lead McGonigle and Chalmers (1997b) to claim that the monkeys could:

> '.....interiorise linear ordering devices which enabled the 'skipping of intermediaries' in five term series problems - based upon transitivity of choice and strong serial position effects in acquisition - and later by a reaction-time based SDE.'
>
> (McGonigle and Chalmers , 1997b: p. 8)

However, in the same chapter, they go on further to cast doubts upon the reasons then held to make so strong a claim. Following a series of post-tests on both their own work and that of Trabasso et. al., (1975), strong directional effects were found, suggesting the existence of 'end-anchor' effects, in which individual subject data revealed 'privileged' salience of the end items (A, or E, in the five-item series ABCDE), (see also Potts, 1972, DeSoto et.al., 1965). The suggestion now was that the SDEs found might not necessarily result from the subject's representing the components of a pairwise comparison as being a subset of those comprising a 'known'

linear sequence. McGonigle and Chalmers had shown that the same result could be derived merely as a function of their immediate 'end-anchor' relatedness alone rather than by some extrapolated ordinal separation. What resulted from this latter analysis was a new position statement for our understanding of linear representation construction. Whereas the transitive data for both the binary and triadic conditions remain consistent, what at first pass appeared to be the construction of a representational structure in which stimuli could in principle be ranked, later appeared to be a much weaker ranking device with strongly unidirectional properties.

Another set of findings indicating that choice transitivity expressed under binary choice conditions was not as complex as originally envisaged (see Trabasso et. al., 1975, for contrastive case) came from the production modelling work of Harris and McGonigle (1994). In their paper, which attempts to model each subject's choice decisions for both binary and triadic phases of the five-item series studies of McGonigle and Chalmers, (1986, 1993), it was shown that only a small subset of rule stacks (sixteen in all) could represent the adjacent pairs of the five-term series (the total permutative set numbering a potential one thousand, nine hundred and twenty rules). Furthermore, all stacks that performed correctly on the initial training pairs also performed correctly on the remote pairs without the addition of subsidiary assumptions or procedures being nested. The model of Harris and McGonigle (1994) thus claimed transitivity to be an integral property of this formal representation. Taken with the above, this new position supported the characterisation of transitive choice as a serial process, but also pointed to the need for more expansive decision spaces to be employed in future studies, the processes involved remaining largely underdetermined by the use of binary tests alone.

Although both the dyadic and triadic transitive data had produced robust search procedures and a coherent story for the evaluation of the dynamics of serial search strategy development, the issues of determining the optimal sequence characteristics, levels of task complexity and increasing difficulty remained to be addressed. A major need at this point was to be in a position that allowed the subject (be it human or non-human) to demonstrate explicit seriation, operating with a large enough array for simultaneous search, rather than merely relying on inference decisions

based upon trained triadic and dyadic-paired choices. Given both the characteristics of the formal model of Harris and McGonigle (1994) and the measures of economy developed by the subject for the making of transitive choice (McGonigle and Chalmers, 1992, 1996), such findings led to what McGonigle and Chalmers called an 'ontological paradox':

> '....that children who as young as four were claimed to reason transitively without direct perceptual support, could not arrange objects in order of size when these were in full view of the subject'
>
> McGonigle & Chalmers (1996 p.13)

The task being referred to here was the now classical ten-item monotonic seriation task developed by Piaget and Szeminska (see Inhelder and Piaget, 1964), used commonly as an index of human cognitive growth. Typically remaining unsolved by human subjects below the age of six years, this task was assuredly more complex than that requiring merely a single binary decision to be made.

Arbitrary and non-arbitrary connectives

Another effect of the move towards a larger set size carrying clear monotonic relations, was that it leant itself to comparisons with sets of equivalent size but with more arbitrary connectives (such as colour orderings), so simultaneously controlling for string length *per se*. Providing a further measure of effective search constraint, McGonigle and Chalmers (1992, 1996, 1997a & b) showed their subjects to be differentially responding to the design primitives employed. Following the claims of McGonigle and Jones (1978), when there was a relational aspect to the successor items to be seriated, a relatively economic search path was produced. For example, it was repeatedly found that for serial monotonic size ordering (as opposed to non-monotonic or arbitrarily associated colour ordering), task completion was by far the most economic, as determined by both acquisition measures and error profiles (McGonigle and Chalmers, 1992, 1997b). Using the same subject within each condition, the development of such tasks afforded the subject the opportunity to demonstrate performance with a view to producing a better characterisation of the progressively economic and self-regulatory hallmarks of cognitive development. Perhaps more difficult to envisage

previously was the solution to the problem of establishing an objective metric of task difficulty, whilst identifying the regulators of growth over the subject's ongoing stages of development. In selecting a serial order problem (requiring at least some degree of on-line executive control), the approach of McGonigle and Chalmers not only afforded a high degree of ecological validity, but also provided a novel task scenario, the very nature of which gave rise to an inherent source of variation in difficulty. This difficulty was based principally on the size of the search space required to solve a particular problem: the greater the search space, the greater the constraint demanded of the agent to search the space efficiently. Two consequences follow from this: firstly, an objective measure of task difficulty was derived from the means of quantifying the search space. Secondly, the cognitive hierarchy exhibited by a given agent would reside in the type and efficiency of the constraint devices used in these situations. In short, the *utility* of cognitive organisation, and indeed its very rationale (see McGonigle and Chalmers, 1997a) was indicated by this argument on its relative success in constraining search effectively.

Establishing an objective task hierarchy : the combinatorial explosion of possibilities

Following this line of work, it is evident that a radically different metric has now been established, based not merely on the *number* of items in the search space (which give a simple interval metric whereby six items differs from five in terms of one item only), but instead is based on the geometric expansion of the search procedures which may be implicated when items increase in what would at first appear to be simple ways. By this method, as the number of items to be ordered increases, so too does the number of possible sequence pathways which might lead to an exhaustive search of the set. Especially in cases for which the task requirement is to provide a unique ordering of the set, the solution is made exponentially difficult and thus raises the stakes combinatorially with each additional item (see figure 3.2 below). Subsequently, if the degree of strain imposed upon the subject is going to be modulated by the length of the sequence to be seriated, then for a longer sequence it would be beneficial to the subject if s/he be able to take advantage of any salient sub-divisions it might allow. Such sub-divisions might suggest themselves in a number of ways. For example, by spatio-temporal displacement or categorical feature

detection, and, if nested in some consistent way, the sophisticated subject might utilise a form of hierarchical expertise in managing the sequence. (In the absence of such a system, of course, only relatively small sequence lengths commensurate with Miller's (1956) infamous seven (plus or minus two) items might be predictive of the upper limits of successful sequencing - even for the relatively expert human adult - due to memorial factors, if unrehearsed or otherwise not belonging to saliently privileged sets).

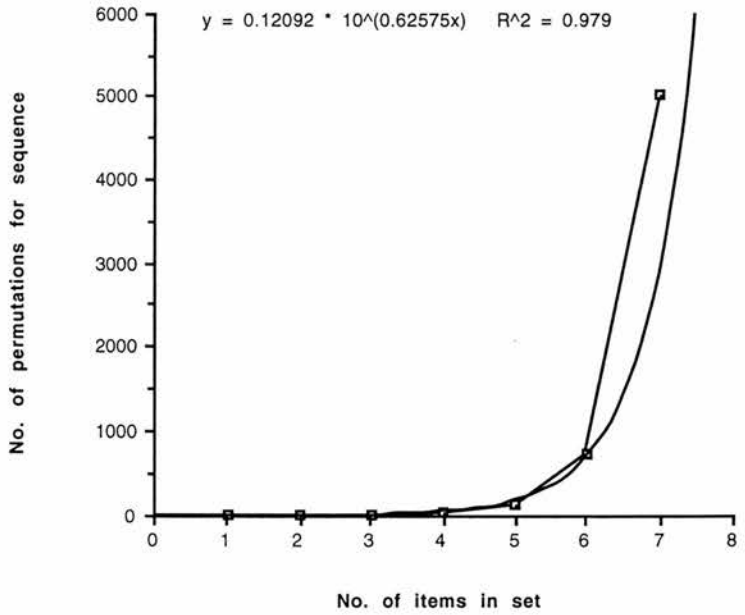


Fig. 3.2 Expanding possibilities with increasing set size.

Cognitive organisation as constraint satisfaction

Confronted with this combinatorial problem, the agent has either to use brute-force memory or devise a means of reducing the cognitive load or strain. Here the role of relational connectives is seen to be crucial (see Terrace and McGonigle, 1994) and, as McGonigle and Chalmers (1992) have argued, transitivity, seriation, and hierarchical organisation are viewed in this context as relationally-based means by which the agent comes to better control their search. This is because relational connectives as exhibited in, say, a size seriation task, can provide simple, data reducing consequences for the agent when it is engaged with a monotonic sequence. This could merely require the iteration of the same relational

rule (e.g., bigger than, or smaller than) and enable the agent to better generate and predict the successor items following the successful seriation of a few antecedent elements from a given series. Over the last ten years of their work, McGonigle and Chalmers (1996, 1997a & b) have provided good evidence for the existence of data reduction strategies and the use of privileged pathways through relatively large search spaces, demonstrating the use of linear and hierarchical structures as efficient, economical strategies for data management in both human and non-human primates. However, given that one still does not know what the core characteristics of human cognition might be which give rise to hierarchical management, taking the stance of McGonigle and Chalmers, and using their proposed behavioural indicants as the existence-proof of higher cognitive functioning in both human and non-human primates, a good candidate procedure is one which builds upon their work on the decomposition of linear seriation and transitivity skills and moves towards an investigation of the hierarchical control and management of serial order production. The animal work reviewed in the earlier chapters had already shown that the separation of a collection of items into groups can be seen as a precursor to the more hierarchical clasification systems which underpin human symbolic processes, and that clasification could be revealed without the use of words. So, using a non-linguistic animal, bereft of any semiotic instruments, it could be hoped that such a study as that proposed by McGonigle and his co-workers would reveal the existence of any emergent behaviours which might support the growth and development of such cognitive organisation.

Linear and hierarchical classification: a new 'voyage of discovery'

In order to achieve a finer characterisation of serial search as discussed above, a new program of experiments (McGonigle & Chalmers, 1996) was embarked upon which was to employ a longitudinal design, following the same subjects over a number of years, using an identical set of procedures and paradigms with both human and non-human subjects alike for comparative study. The non-human component of this larger study of McGonigle involved the use of brown capuchin monkeys (*Cebus apella)* whose development over a five year period make up the core of the experimental work to be reported in this thesis. Situated in the Laboratory of Cognitive Neuroscience and Intelligent systems at the University of

Edinburgh, and, in accordance with the suggestions and recommendations for both husbandry and experimental environments outlined above, a new hierarchical classification paradigm of McGonigle addressed the question of whether the non-human subject(s) might be capable of exhaustively searching a set-size of sufficient length to then allow its subsequent decomposition into categories within which individual exemplars themselves then be further explicitly ordered (see also McGonigle and Chalmers, 1996). One can imagine from the outset that in order to do this it would be necessary to operate with a minimum set size of at least four items, providing two categories, each containing two differentiable exemplars. From the literature to date, one would expect both human and monkey to achieve this minimum number for simultaneous search, D'Amato and Colombo (1989) having previously reported serial learning results for a monkey working with a five-item set. However, there was no indication from the literature as to what kinds of sequential constructions the monkey might be capable of in a non-tutored task, nor was there any hint as to the limitations which might be imposed upon the number of categories or categorical exemplars to be employed in a program of this sort. It was with this born in mind, therefore, that the ongoing stages of the hierarchical classification series of experiments set out on a 'voyage of discovery', taking six adult *Cebus apella* first through a supervised learning procedure using tasks of increasing complexity and levels of difficulty, in the laboratory. At each stage of the experimental series, the inclusion of a particular individual would be determined by its own success, itself being continually informed by the progressively developing expertise acquired from the prior 'easier' tasks as later determined once the level of difficulty had been raised. Once again, it was important to emphasise at this point that there were no *a priori* indicants suggesting that the Cebus monkey was going to be able to succeed with a set size anywhere near a sufficient length for seriation of the sort comparable to the Piagetian ten-item monotonic series mentioned in the previous section. For example, if a non-linguistic agent was to be able to successfully report searching a sequence composed of nine linear (size) items arranged in a monotonically ascending order, the solution requires the subject to predict successor items on-line, whilst maintaining good memorial control of sequence position in real time. Without establishing at least some prospective criteria for an 'optimal' minimum string length for subsequent size-relational interval discrimination at a given level of

cognitive development, if the monkey (or young human child) were to fail at such a task, how would such failure best be interpreted ? Using the established and robust procedures for serial exhaustive search already employed with non-human primates and the tried techniques with touch-screen apparatus (De Lillo, 1994; McGonigle and Chalmers 1993), this relatively open exploration was to be implemented according to the following design characteristics with a view to assessing the qualification of (if not actually producing) a set of candidate non-human subjects for the more liberated, free-choice, serial search demonstrations typical of the more sophisticated seven year old human child or adult. In the more recent experimental work to be reported below, both the apparatus and the species was to have changed, but remained situated within the laboratory of McGonigle at the University of Edinburgh.

A general rationale for a comparative classification and hierarchical organisation study

Now based on a thriving colony of Brown Capuchin monkeys, *Cebus apella*, established and settled into a highly enriched colony environment, McGonigle has continued the development of an even wider range of new procedures based on touch-screen technology which has enabled the extension of his original work in directions never before possible using the WGTA. Motivated by their more recent human child developmental work also using touch-screen based techniques (McGonigle and Chalmers, 1993, 1996; Chalmers and McGonigle, 1996) the suggestion was that cognitive agents could demonstrate data reducing strategies such as classification and chunking in achieving cognitive economy in serial ordering tasks. The study to be reported below implements a comparative behaviour-based paradigm using a non-linguistic agent (McGonigle et. al., 1994) in a first attempt to compliment their human child work on classification. Entirely success-based, each individual monkey was to be presented the opportunity to move through a series of increasingly demanding tasks, the level of difficulty starting with simple relational discriminations through conditional rule-based learning, then onwards to seriation, transfer and finally to classification, as near as possible procedurally complementing the human child studies (McGonigle and Chalmers, 1993, 1996; Chalmers and McGonigle, 1996). Using an ordering task (seriation) embedded within a multiple classification procedure, this

paradigm was designed such that one might for the first time be in a position to assess and reveal hierarchical cognitive organisation without the use of linguistic tasks. The general aspects of the study will now be outlined prior to the presentation of a more detailed and specific exposition of the procedures to be implemented in the following chapters.

In order to be in a position to reveal such a hierarchical organisation should they be constructable, the monkeys were initially and explicitly trained to order icons (on the touch screen) drawn from putatively different categories composed of coloured, simple geometrical shapes: e.g., [A] red star, [B] yellow hexagon, [C] blue triangle, etc., arranged within spatial arrays which would vary randomly from trial to trial (McGonigle et. al., 1994). In accordance with the schematic shell of the tasks shown in figure 3.3, the behaviour-based paradigm was designed such that following acquisition and consistent performance with, say, a three-item ordered list [ABC] the sequence length could then be extended in either one of two main ways. The first, (increasing breadth) would be to add novel stimuli from new categories, providing a learning situation akin to learning an alphabet [ABCDE...]. The second, (increasing depth) would be to again increase the sequence length, but this time introducing

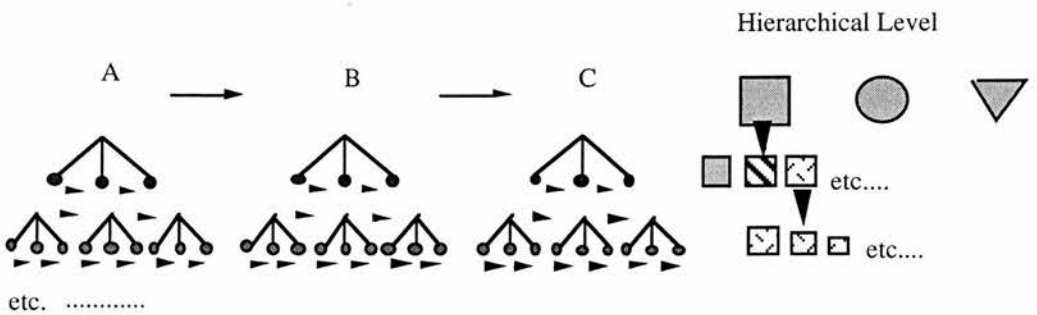


Fig. 3.3 Hierarchically-based search trees as implemented on the touch-screen for both human and non-human primates. (after McGonigle and Chalmers, 1997b, reproduced with permission).

opportunities for classification *within* the existing string, say [AABBCC]. Under these conditions, a free choice stage would be provided. The object here would be to ascertain whether the subject might spontaneously classify by reporting all physically identical items of the same ordinal position prior to reporting the items of the next (e.g., all of the [A]

exemplars before all of the [B] prior to those of the [C] ..., and so forth). Failure to classify, seen as the inability to take advantage of such data compressing possibilities, would result in strings of disorganised sequences being produced such as ABBACA. The limits placed upon the string lengths that could be controlled (without errors) would be similar to those recorded under single exemplar conditions, the latter only being learned by the use of brute-force memory or Ebbinghaus-style rehearsal. However, if classification and chunking was going to be used by the subject, more extended and strategically (possibly hierarchically) organised strings such as A1 A2 A3 B1 B2 B3... etc., might emerge.

It was also known that a further index of successful classification might be based upon any phrasing effects to be found, as had been discovered in a number of child studies. Although it might not be possible that the monkey be capable of sustaining consistent and uninterrupted engagement with the task for the duration of a single trial, it may be seen from figure 3. 4, that such pausing at the categorical boundaries might occur. Here the search of items within a class being generally much faster, once the first exemplar of each category had been located (McGonigle, 1987; McGonigle and Jaswal, 1993).

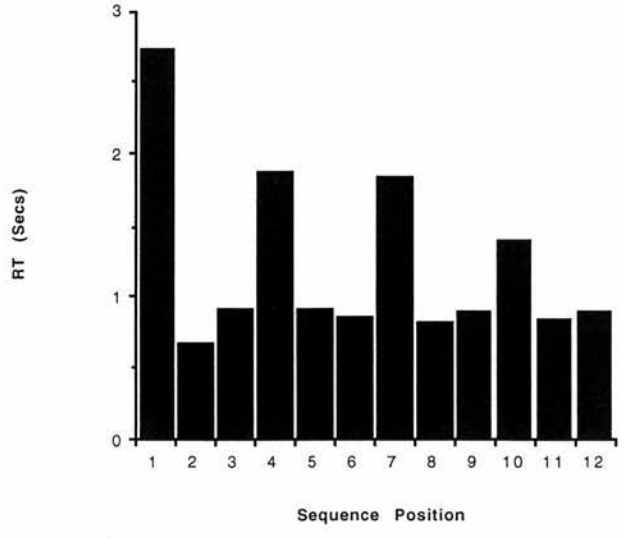


Figure 3.4 Mean reaction time (RT) profile for successful 12-item sequencing of four categories, each with three exemplars by a human child (aged 5 years)

However, simple classification of this sort would not be effective over very long sequences when the category size became too large, resulting in a

search problem of its own. A solution to this problem would be to achieve some form of hierarchical organisation which could keep the search within manageable proportions. If sequences of sufficient length were found to be sustainable by the monkey, with the subsequent introduction of physical variation amongst the exemplars of each class one could assess this competence further by imposing *within*-class ordering requirements. For the human child, exemplars have already been subject to variation in both size and colour, and this for the monkey, would enable us to determine the levels of equivalence which they might adopt when classifying. Given any degree of success, and this was by no means preditable at the time, it might then be possible to demand the ordering of each of the individual exemplars within each category as well.

Given that the extant literature offers no insight as to the possible number of classes and/or exemplars of classes one might expect to see the monkey to comfortably operate with, using an incremental procedure, the number of items in a sequence to be reported was to be gradually increased according to an individual monkey's levels of continuing success with subsequently larger sequences within an experimental series involving increasingly complex tasks. Given that this was to be an open-ended experiment, an initial choice of three categories, an [ABC] core sequence was assigned to each monkey and, if achievable, would thereafter form the basis for further extensions by breadth and/or depth.

Given a degree of success with the production of sequences which might prove to be extendible, there were a number of ways that such extendibility might be implemented. For example, individual icons might be made salient by their differential rotation about some axis by varying degrees, or they might be free to vary by some other morphological feature. It was already clear from the prior work of McGonigle and Chalmers, that much to be gained from the use of both monotonic/non-monotonic and relational/arbitrary iconic connective distinctions, and so it would seem in keeping with their findings that the most appropriate icons to be used for the purposes of further investigating hierarchical organisation by classificatory processes would include simple geometric forms free to vary in size and/or colour.

Now in a position to embark upon on our 'voyage of discovery', many of the problems concerned with the shortcomings of previous attempts to reveal the regulators of cognitive growth have been addressed. Given that our provision of an optimal environment for the study of comparative animal cognition has now been better chracterised, it is with respect to those needs outlined above that a longitudinal program of studies was to be embarked upon. Working within the new agenda set by McGonigle and his colleagues, it was only now, and for the first time, possible within a single paradigm, to work with novel behaviour-based methods which neither require (nor assume) the existence of linguistic abilities in the subject. Such a method also allows for the evaluation of a species comparative study using a non-linguistic task with subjects devoid of 'semiotic instruments'. The identification of design primitives has clearly been made from which more complex cognitive skills might be expected to derive, using procedures which were designed to fractionate such 'higher' cognitive processes into their more basic component building blocks. Using a serial search task, entirely success-based (and by no means success-assured), each individual monkey was to be presented the opportunity to move through a series of increasingly demanding tasks, so assuring the continued presence of a degree of cognitive strain in need of some adaptive, on-line control on the part of the subject. The actual string lengths and sequence compositions implemented were drawn from a large number of possible decompositions available for the expansion of a simple [ABC] icon set, the background conditions requiring antecedent successes so as to provoke continual monitoring and ongoing descriptive analyses in the subjects as they proceeded with prospectively larger search spaces.

If any of the *apella* subjects were to reach a level of expertise, say, with at least three items in each of at least two categories, once the hierarchical phase of the study was underway, it would quickly become evident whether the monkey could use class-based structure to help manage long sequences.

The first of these measures would be simply describable in terms of the length of sequences which these subjects might control when classification had become an established option. The second would be in observing the degree of spontaneity with which subjects searched for items that were physically similar or resembled one another, rather than

continued to select items in a different category. Third, there would be an overall longitudinal pattern emerging from the life-history of each subject in the ongoing programme (becoming established over a number of years rather than months) indicating the degree to which an individual monkey might be coping with progressively more difficult problems in the face of reducing material rewards, with less cognitive cost. As the tasks could be objectively measured in terms of their increasing difficulty by computing their combinatorial expansion with each increase in sequence length, *ceteris paribus*, the strategic benefits which classification might appear to provide as a data reducing strategy could always be indicated by accelerating acquisition rates (McGonigle and Chalmers, 1996). Indeed, outside the possibility of any chunking or other grouping of items together, the degree to which a monkey's exploitation of a simple classificatory scheme might be to its advantage, would be made transparent when their acquisition performance for varying sequence lengths be plotted against the combinatorial expansion for which the management of such sequence length increases otherwise entailed.

If there be any aspect of a cognitive mechanism that sets itself apart from a more reflexive mechanism of learning and adaptation, it is likely to be discovered by inquiring of the functional role of cognitive processes. Whatever the benchmarks of success to be achieved by any individual subject in the course of their journey through this open-ended series of experiments, the vista ahead, informed by progressively deconstrained future free-search conditions, would help determine whether any indicants of the deeper cognitive structures central to human cognition, could be found in the non-human primate. If they were to so emerge, possibly epigenetically, they will have done so as a result of their ongoing interaction with high cost ordering tasks for which classification might serve an important strategic, information management, data reducing function.

Chapter 4

Classification and Hierarchical Organisation in *Cebus apella:* The study

Part 1: Classification and control of extended serial order productions.

GENERAL PROCEDURES

All of the experimental work to be reported here was conducted at the University of Edinburgh, Scotland, within the Laboratory for Cognitive Neuroscience and Intelligent Systems under the supervision of its founder and director, Dr. Brendan McGonigle. A thriving colony of Brown Capuchin monkeys, *Cebus apella*, have become established and settled into a highly enriched colony environment, providing a new facility which was specifically designed to provide the most optimal conditions for the promotion of sustained monkey performance throughout longitudinal studies (to span across years rather than months). At the same time, such an environment affords the continued development and expression of any cognitive competences which they might reveal themselves to possess.

In a preliminary report by McGonigle and Jaswal (1993), spontaneous classification was seen to be produced by young nursery children with a serial search task. In the case of the human study, the subjects ably demonstrated an [ABC]- core sequence extended to include a fifteen-item sequence (five categories with three identical exemplars of each), but also showed a consistent RT phrasing pattern related to the category boundaries (see fig. 3.4). Furthermore, not only did the individual latencies rise at every category boundary during sequence execution, as the subject progressed down the sequence, the time taken to touch the first icon of each new category became shorter although there were no significant changes in the intra-category latencies throughout the sequential production. This finding was interpreted to suggest that the subject was not just making pauses, but that once the first item had been identified, then the subsequent items to be searched required relatively less effort as the sequencing continued.

Prior to our attempt to replicate this study with a non-human primate,

DeLillo (1994) had already established the necessary preconditions for the *Cebus apella* to enter the stage of the research program to be reported below, having conducted a study with simultaneous exhaustive search tasks with the same monkey subjects. In these tasks the monkeys were incrementally exposed to a maximum of nine physically identical icons in multiple locations on a touch-screen for which a serial free-search solution was required. The procedure employed at first did not penalise for reiterative touching, and so afforded the subject the possibility of developing strategic devices in order to reduce the search space and the consequent effort required to negotiate it. The results indicated that the monkeys were indeed capable of self-regulating their responses (well beyond the task demands or machine requirements) in such a way that they came to determine, and execute, more economic pathways through the space prior to exhaustive search success. These strategic devices included the use of preferred starting positions, adjacency movements and privileged vectors. For these free-search conditions, DeLillo (1994) also reports significant increases in non-redundant touching of icons as being further indicative of self-regulation as motivated by a need for cognitive economy (McGonigle and Chalmers, 1992). In a short series of experiments involving classificatory options (still under free search conditions) there were only weak indications for the use of any spontaneous classification of like icons with distinguishing colour features arranged in either rows or columns (DeLillo, 1994). What was, however, being provided by this pioneering set of studies, was a warrant for the experimental design and further elaboration of McGonigle's paradigm for further implementation. This and other more comparative studies (e.g., McGonigle and Chalmers, 1980) offered good grounds for thinking that such non-trivial tasks could provide a useful measure by which to evaluate both the quantitative and qualitative differences between species (at that time specifically referring to human and non-human animals).

The following section provides a brief history of the monkey colony adults participating in the study, together with the details of a novel supportive laboratory design, its husbandry prerequisites and specifications.

The Cebus apella colony and laboratory environment used in the study.

The *apella* monkey colony derives from an initial group of six feral born sub-adults (two male and four female) estimated from body weight, colour and dentition to have been between three and four years of age upon their arrival at the laboratory. They were transported directly to the Laboratory from South America in 1988 specifically for the purpose of establishing an in-house breeding facility for longitudinal research work in comparative and developmental cognition. Following an immediate six-month period of strict veterinary quarantine, the females were quick to conceive and have to date produced a further thirteen healthy youngsters, all of whom have survived. One advantage of the establishment of such a colony is that many of the otherwise confounding environmental variables which might contribute to the interpretation of any behaviour changes recorded in an animal's development would be known. Such a history has rarely been compiled for many laboratory primates and quite frequently the detailed hereditary, social and circumstantial background of an individual animal subject will remain unknown or incomplete. Health records are more widely and routinely kept, but by themselves do not offer sufficiently detailed information which might otherwise assist in the explanation of the differential performance characteristics often revealed in the course of laboratory experiments.

As long ago as the 1930s, Heinrich Klüver accorded his continuing experimental (and somewhat unusual breeding) successes with New World primates to be largely the result of his day-to-day husbandry and dietary practices (Klüver, 1933). As with the case of the human child, one should not expect isolated subjects living in sterile, featureless environments to be capable of demonstrating any normative developmental processes of any kind, let alone to reveal the direct effects of some experimental variable to be assigned at some later point by the researcher. For these reasons (although by no means an exhaustive list) every member of the monkey colony resident in our laboratory enjoy what we would propose to comprise the most optimal conditions for their continuing high levels of individual physical and mental health, collective social welfare and comfort. Such attention to the husbandry environment has, I believe, paid itself dividend as seen by the high degree of co-operation seen with our monkeys; their voluntarily working on a daily basis, whilst consistently

showing attention and orientation to their task environments. Most important of all, changes reflecting the effect of traumas (resulting from deteriorating health, social withdrawal or lack of stimulation, for example) which would otherwise necessarily confound any explanations of the many changes recorded during an animal's behavioural development, cannot go undetected. This is of special relevance to those instances which might involve the interpretation of experimental performance stasis or failure with a novel task, typically resulting in much reduced numbers of monkey subjects being reported upon than were actually used in studies of animal learning in the literature.

Each of the current total of nineteen monkeys live in a family group enclosure within one of three colony rooms, all adjacent to, or nearby their daily experimental testing-room (see figure 4.0 for layout and dimensions). Every monkey has continuous conspecific tactile stimulation opportunities, and each enjoys auditory and olfactory contact with at least five other members of the colony, making visual contact with as many as a further nine others at any given time. Grooming and group play are common throughout the daylight hours and at no time is any animal housed alone, excepting rare incidences of veterinary attention being required. All monkeys (with their own captive-bred and laboratory-born offspring) are free to forage and play together both day and night in full-volume densely branched and activity decorated enclosures with permanent deep-litter flooring and 'off-exhibit' private areas. Plates 4.1-4.3 show typical colony rooms of the Laboratory and some of the environmental features installed in each (natural branches at various heights, rope-swings, deep-litter forage and various 'hides'). The colony rooms are serviced by one part-time and two full-time animal technicians seven days each week, providing each room with at least a weekly all-surface wash and daily shelf and floor-litter cleans as required. Unlike so many animal learning laboratories in the past, no food deprivation or weight-control is imposed upon the experimental animals prior to their daily experimental sessions. We have found instead that the use of a daily, balanced-diet feeding schedule (Mizuri nut staple supplemented with fresh fruit and vegetables) in no way interferes with the monkey's motivation for experimental reward for up to a maximum of fifty single peanut deliveries (a preferred food) in a single daily session. Irrespective of the level of an animal's performance, at the end of week each monkey

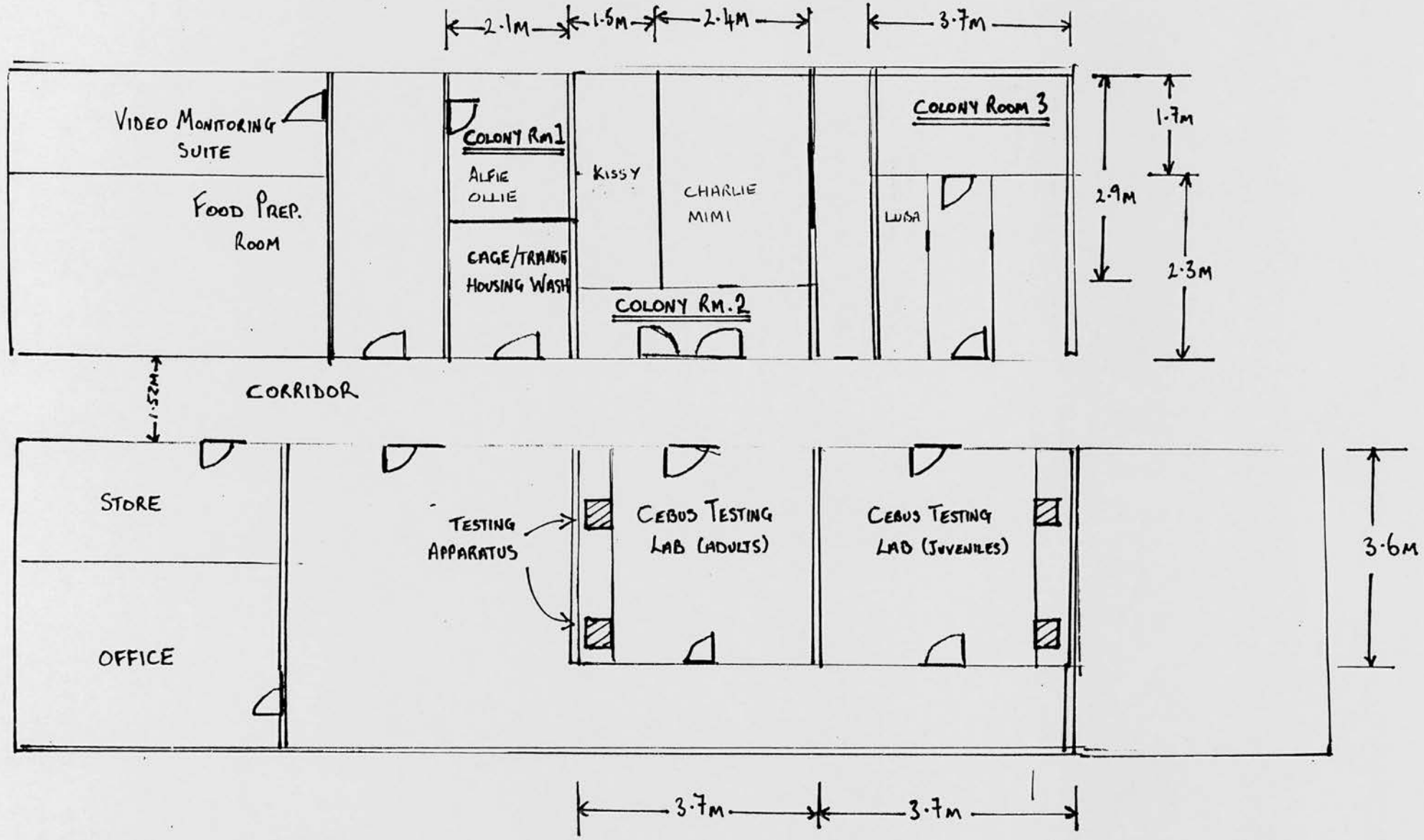

2·1M
1·5M
2·4M
3·7M
VIDEO MONITORING SUITE
FOOD PREP. ROOM
COLONY Rm1
ALFIE
OLLIE
KISSY
CHARLIE
MIMI
CAGE/TRANSF
HOUSING WASH
COLONY RM.2
COLONY ROOM 3
LUISA
2·9M
1·7M
2·3M
1·5M
CORRIDOR
STORE
OFFICE
TESTING APPARATUS
CEBUS TESTING LAB (ADULTS)
CEBUS TESTING LAB (JUVENILES)
3·6M
3·7M
3·7M
= ANIMAL HOUSING

Plate 4.1 Two of the *Cebus* family colony rooms at the Laboratory for Cognitive Neuroscience and Intelligent Systems, University of Edinburgh.

Plate 4.2 One of six deep litter, enriched environment in a typical *Cebus* colony room.

Plate 4.3 One of the Cebus subjects at play in home colony room.

receives extra fresh fruit forage, mineral and vitamin supplements. Other, less frequent 'treats' might include the provision of whole nuts, hard-boiled eggs, chillies and ice-cubes (these latter two items appearing to be treated as extra 'toys' as much as foodstuffs !!). Water is provided *ad libitum* and monitored throughout the day for freshness.

Regular veterinary inspections are carried out for all monkeys, during which time their dentition and manicure needs are attended to. This is the only time that the monkeys are directly handled or held, although tame to all of the Laboratory staff. For the purposes of their daily transfer to the experimental test-room (and for weekly weighing), each monkey will voluntarily withdraw from his/her home colony room directly into a wheeled transit-enclosure. Plates 4.5 - 6 show the movement sequence of a monkey from 'home' to the experimental testing-room. In keeping with our Laboratory philosophy of optimising performance, the monkeys work at their tasks simultaneously in pairs, side by side, in a manner related to their more natural and social foraging habits. The females will typically work with their nursing offspring when they have them and auditory contact with the colony groups just left behind are always maintained during the experimental session. After the animal's experimental session has ended (a maximum period of one hour), selective attention may be given to the needs of any individual monkey prior to their otherwise immediate return to their colony room. Indeed, apart from their required attendance at the daily experimental session, a monkey is very rarely withdrawn from its family environment.

The conditions of our primate section of the Laboratory for Cognitive Neuroscience and Intelligent Systems as outlined above, has produced a flourishing colony of *Cebus apella* monkeys, optimally reared for their participation in a longitudinal study within a larger program of developmental and comparative animal cognition. Having thus designed, observed and maintained improvements in diet, housing enrichment, socialisation opportunities and health management, we have successfully reared animals in the absence of any schedule-induced stereotypies, who clearly demonstrate consistent motivation suitable for their inclusion into the working environment of the developmental programme now to be described.

Plate 4.5 Experimental test laboratory showing *Cebus* transit/working housing and touch-screens.

Plate 4.6 Cebus monkey working at touch-screen with automated peanut dispenser in foreground.

Simplified design and method.

Icon sequencing characteristics and reaction time data were recorded for subjects working through acquisition, classification and explicit seriation of up to 12 item arrays presented on a computer touch-screen apparatus. A total of 45 consecutive phases have comprised the principle experimental series of the study to date. Individuals self-qualified by success for inclusion in subsequent phases of the experiment using a longitudinal within-subject design.

Subjects

Six sexually mature adult *Cebus apella* monkeys participated in the study, two male, four female. All of the monkeys had prior experience with the apparatus, previously engaged in a similar experimental protocol (De Lillo, 1994; McGonigle and Chalmers 1993), but this was to be their first exposure with multiple icons of more than one type being simultaneously shown on the touch screen. Table 4.1 shows some demographic details of the monkey subjects participating in the study, with ages given as being true at their time of entry into the experimental series to be described below. All of the adult monkeys were feral born and caught, and have been subsequently responsible for parenting the juvenile offspring family-reared with them in the laboratory's open colony rooms. Monkeys were only housed individually for the duration of their experimental session, and the involvement of any one subject at a particular phase of this experimental series (through a possible total of between 1-45 phases) was determined by the progressive performance assessment of each.

Table 4.1. Cebus apella monkeys participating in study.

Monkey Name	Gender	Est. Age	Origin
Charlie	Male	9-10 Yrs.	Feral born
Alfie	Male	9-10 Yrs.	Feral born
Mimi	Female	9-10 Yrs.	Feral born
Luba	Female	9-10 Yrs.	Feral born
Olie	Female	9-10 Yrs.	Feral born
Kissy	Female	9-10 Yrs.	Feral born

Stimuli

As stimulus material, up to four each of three coloured simple geometric figures (occupying a 35 x 40 mm area) could be randomly positioned and presented within a symmetrical five by four matrix (170 x 190 mm) on the surface of an interactive video touch screen. Pixel densities were constant for each target and were generated by a computer graphics system, the software for which was generated in-house using the C-programming language. Each subject was assigned a unique six-item category sets [ABC and DEF] of differing shape and colour (e.g. red square [A], green circle [B] & blue triangle [C]) so controlling for stimulus-preference effects across animals. For use in the later equivalence phases, each subject's stimuli pool contained three coloured, two monochrome and nine sizes of each icon.

Apparatus

The experimental test environment used for all phases of this study may be seen in Plates 4.5 - 6. Subjects were situated in a wheeled transit enclosure (610 mm x 880 mm x 610 mm) with one side opening within easy reach of a 14" Colour (VGA) interactive *Micro touch*® touch-screen monitor. The bench height above the floor was fixed at 830 mm with the screen proximity adjustable to the monkey's required comfort when operating the system. Using in-house purpose-written software, test stimuli were generated and presented by an IBM-compatible 386 PC which also supported automatic recording of the touch-screen response data in real time. The experimental programme also controlled a stepper-motor drive peanut dispenser mounted to the frontal left-hand side of the working environment. Session touch-screen behaviours were also recorded on video cassette tape for all trials and subsequently archived and edited for off-line analysis. All sessions were conducted within the purpose designed laboratory test-room with controllable lighting levels, also providing a background of 15-20 dB broad-band white noise sound via a signal generator constructed in the departmental workshop.

Pre-training

In each phase of the experiment the monkey was required to interrogate the visual array and to sequentially touch every icon presented on an interactive video touch screen. At no point were the monkeys task-informed other than that they were to initially approach the screen with a forward-facing hand reach, and to touch a single coloured geometric shape appearing at a random location on the touch screen. Reinforcement (free feed) was offered for successive approximations towards icon-touching posture once screen-positioned. The monkeys all had prior experience with the touch-screen environment to be employed, having previously worked on a task involving serial, exhaustive, spatial, free-search experiment employing multiple exemplars of a single icon-type (De Lillo, 1994; McGonigle and Chalmers 1993). No monkey had ever been provided differential feedback for a specific sequence of touches, latency marking or any other performance parameter incidental to serial exhaustive search.

General Procedure.

Having voluntarily withdrawn from its family-grouped colony room and situated in wheeled transit housing, each monkey was tested daily at a familiar fixed location. The monkeys worked in pairs inside the purpose-built testing suite in the laboratory, situated at adjacent testing areas within visual and auditory range of each other. Lighting was subdued so as to provide better focus towards the touch-screen environment and a white-noise generator was operating to filter out extraneous auditory inputs from the nearby monitoring suite and offices. All subjects, once comfortable in the test environment, were presented with a number of icons upon the touch-screen monitor and were required to make serial exhaustive touch responses towards them with an outreached arm. The novel aspect of this particular series of experiments was that serial exhaustive search should continue only after all the icons of a given category had been searched, in a consistently ordered manner, irrespective of the number of icons presented (i.e., touch once, each member of category A before each member of category B, before those of category C). At no time throughout the experimental series was the subject given any explicit tuition, or engaged in other experimenter interaction

(although sometimes present). Having touched a target icon, a tone would sound and the target would be highlighted, then disappear from view for 200 mS. For all screen displays in the experimental series, after each of the targets had been correctly touched, the array was removed from the screen which blanked for 14 seconds. If, and only if, the touch sequence was executed without reiteration of any of the target items to be ordered in the array, a single peanut was delivered by the dispenser. Repeated touching of a given icon prior to moving on to the next was permitted. If, however, a subject showed a reiteration of target touch responses for the current screen array, the screen became blanked white for 30 seconds, the feed-dispenser remaining inactive. The inter-trial interval for correct trials was set at 14 seconds, the whole session capable of automatic completion either following the completion of a given number of trials, or until the reaching of the experimental criterion pre-set for the phase. Whether or not the monkey had successfully completed any given trial, each subsequent trial presented the icons of the same phase's conditions in a new configuration, each to one of the twenty possible positions determined by a pseudo-random number automatically generated by the computer program. Subjects could be taken through sequences of incremental item numbers and combinations of category arrays in a series, up to and including a maximum of 12 items per screen array, dependant upon their degree of success and task-motivation as determined by the monkey's successive phase criteria performances. The entire 45-phase experimental series may be seen as comprising four principal sub-divisions:

(a) an acquisition phase [ABC phases 1-11, DEF phases 15-17]
(b) equivalence phases [ABC phases 12-14, DEF phases 18-20]
(c) serial order phases [ABC 2-level (colour and size) hierarchy phases 24-39]
(d) linear size-seriation phases [ABC 6- and 9-item set seriation phases 40-45]

Phases 21-22 afforded the chance for the monkeys to freely interrogate a combined, simultaneous [ABC+DEF] six item array. Phase 23 involved just one monkey (Charlie) in a set of probe trials for the effects of selective loading on serial positions within an extended six item set [ABCDEF].

A subject would receive remedial trials as appropriate after a declining performance over 10 sessions, and this would typically involve a return to

the previous phase completed, once again to satisfy the criterion level of performance before continuing with the next phase once more. At no time was differential feedback offered during the entire experimental series. Each Cebus was run within the same subject-pairing in a pre-determined and fixed order, for a maximum or 40 minutes at the same hour of the morning each day, 5 days a week. The experiments were automated for both stimulus presentation and feeder access components in daily sessions, each containing a maximum of 50 trials per subject per day unless a phase criterion was met within the session.

In accordance with the rationale and figure 3.3 above, the following Tables 4.2 - 48 show the individual phase results and qualifying conditions for the participation of each monkey in the consecutive experimental phases 1 through 44 to date (together with their interspersed conditions and completion criteria). These were to be completed by each monkey in accordance with the specific phase procedures as listed below, each in the order in which they are presented. Although six animals entered the experimental series from the outset, it was not possible to predict in advance how many of the monkeys might succeed at any given level of difficulty as the sequences to be interrogated became longer and more complex. Because each animal would qualify for its inclusion in a particular phase only having reached criteria level performance at the previous phase, descriptive results including the numbers of trials, errors and percentage errors to criteria will be given for each phase and sub-condition. By so doing, it will quickly become evident how each monkey's qualifications for continuation through the experimental series were developed, both longitudinally and with respect to the performance of the other monkeys participating in the study. As a general code, the following key provides examples of the nomenclature to be used in identifying string composition throughout the following tables (see also Appendix x for full stimuli sets [A-F] as actually prepared for each monkey):

A= 1st category coloured shape (e.g., red star)
B= 2nd category coloured shape (e.g., yellow hexagon)
C= 3rd category coloured shape (e.g., blue square)
A_C= 1st category shape free to vary in colour (e.g., red, white or blue star)
A_S= 1st category shape free to vary in size (small, medium or large star)
A_{C1}= 1st category shape, 1st colour in serial order position
A_{S3}= 1st category shape, 3rd size in serial order position

Specific phase details and procedures.

(a) [ABC] Acquisition Phases

Introduction

The initial [ABC set] acquisition phases 1 through 11 were given in the order as shown in Table 4.2 below and include the details of interspersed conditions and completion criteria for each phase. All of the six monkeys participating in this experimental series started with Phase 1 [AB]. All phases required an extremely high level of vigilance and unusually consistent performance in the face of increasing task difficulty. For the subsequent phases, individual monkeys were self-selecting by their continuing successful attainment of each given phase criteria as they progressed through tasks of increasing string length and thus greater levels of difficulty.

Table 4.2. Acquisition Phases 1-11 [ABC]: conditions and success criteria

Phase No.	Conditions	Completion criterion
1	AB	90/120
2	AB	15/20
	AAB	15/20
	ABB	15/20
	AABB	15/20
3	ABC	90/120
4 *	ABC	3/4
	AABBCC	-
5	ABC	3/4
	AABB	15/20
	BBCC	15/20
	AACC	15/20
6	ABC	3/4
	AAABBB	15/20
	BBBCCC	15/20
	AAACCC	15/20
	AABBCC	15/20
7 *	AAABBBCCC	-
8	AABBCC	3/4
	AAABBBCCC	15/20
9	AAAABBBCCC	15/20
	AAABBBBCCC	15/20
	AAABBBCCCC	15/20
10 *	AAAABBBBCCCC	-
11	AAABBBCCC	3/4
	AAAABBBBCCCC	15/20

* = Non-criterion transfer phase

The actual sequence compositions presented during these acquisition phases were drawn from a larger number of possible decomposition strings derivable from an [ABC] icon set, and were those conceivably simpler in structure by comparison. Apart from the specific [AB] and [ABC] training phases (Phases 1 and 3 respectively) only category order was required to be reported by the subject for the sequences of this acquisition phase. There was no explicit tutoring of any monkey for single unique solutions in any other conditions. This was designed so as to avoid the production of stereotyped motor responses whilst at the same time provoking continuous array interrogation on the part of the subject from trial to trial. The use of interspersed conditions of varying string length and compositionality was also programmed for this reason, so preventing the subject's responses becoming too reflexive and thereby less 'cognitive' in production. It will be remembered also that the actual icon configurations were free to vary randomly on a trial to trial basis for a particular condition.

At the time or writing this next section, all of the monkeys had reached at least the nine-item acquisition stage (Phase 8) and for each phase of the study to be reported below, separate rationale, design, procedure and results sections will be presented. For ease of clarity and more consistent comparison across phases, the numbers of trials and errors to criterion will be given for each consecutive phase together with selected reaction time (RT) performance data for correct sequential touching of icons in the test conditions. So as to ensure that such RT data may be providing the most reliable indicants of comparative performance, all of the analyses to be presented below include measures from the last twenty correct sequence trials only (close to the criterion run for each phase). Prior to reaching this level of expertise, particularly with the longer string length conditions, more inclusive data sets would otherwise offer RT measures that include eccentric values tending to vary idiosyncratically with each monkey's condition acquisition learning as well as criterial attainment. To provide further consistency and ease of comparison, numbers of errors and percentage error values will be used rather than the numbers of trials (which might vary quite widely between individual monkeys) to criterion. Significance values for individual monkey RT profiles were always calculated according to an $F_{19,20x-1}$ statistic for 1-

way ANOVA (where x = string length) unless otherwise indicated, and related-t_{20} for adjacent pairwise comparisons of choice reaction times for successive item touches. Although more extensive Bonferoni analyses have been conducted for all sequence lengths, rather than provide exhaustive cataloging, significant increases in RT for adjacent sequential touching will be reported as they occur only at the icon category boundaries ($p< 0.05$) unless otherwise stated in the text. Interim summaries will also be provided for phase groupings as determined by landmark achievements, and at points later in the study which allow for a more convincing contrast being made for the same subject's choice RT performance data variation within the same session (in particular, where sequence length remains constant despite variable compositionality).

Phase 1

Rationale. To allow the firm establishment of the core sequence [ABC] for subsequent generalisation and multiplexing, an initial two-item string [AB] was presented for discrimination and ordering. This first phase would inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli in a consistent and orderly manner. For this reason the experimental criteria for success with this foundation phase was very strict, requiring a very high proficiency mark of ninety correct out of the last one hundred and twenty trials (seventy-five per cent correct).

Subjects. All of the six adult monkeys described above took part in this phase.

Design and Procedure. In accordance with the general procedure above, Phase 1 presented a single condition [AB] with all subjects using two icons discriminable by both colour and shape shown simultaneously to the touch-screen. These were to be interrogated and touched sequentially - [A] first, followed by [B]. Only one error-type was possible (touching icon [B] first), repeated touches to [A] being permissible. The very strict phase criterion was reached when ninety out of the last (cumulative) one hundred and twenty trials had been correctly completed.

Results. All six monkeys completed Phase 1, achieving the level of performance as set by the experimental criterion for continuation with

the experimental series. The number of trials, errors and percentage errors to criteria are shown in Table 4.3.

Table 4.3. Nos. of trials, errors and % errors to criterion [AB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	165	60	36.4%
Alfie	386	169	43.8%
Kissy	235	93	39.6%
Mimi	246	108	44.0%
Luba	169	60	35.5%
Ollie	313	146	46.6%

Discussion. The successful performance of each monkey in this first phase has shown that the choice stimuli for the initial two category shapes were readily discriminable and consistently orderable at a sustainable level over several daily sessions. However, given that only one error type was possible this is not an achievement requiring much explanation beyond simple binary discrimination ability. Given the success of the Cebus on this initial training phase, rather than immediately adding a third category it was desirable to see whether this two-item series might be extended in length following the addition of further icons of the same physical appearance.

Phase 2

Rationale. Having established a two-item [AB] sequence for each monkey, there is at this stage no warrant for the belief that the addition of a third category would be readily taken up in such a way as to generate a three-item ordered set. In this second task, therefore, a generalisation phase would allow us to assess the degree to which a subject will continue to classify additional icons firstly by category, increasing the string length to be interrogated, without uniquely specifying the order requirements to be imposed on the subject for exemplars within each of the existing categories [A] and [B].

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. All subjects were presented with this second phase comprising four conditions [AB, AAB, ABB & AABB], using one or two

identical icon exemplars of two categories shown simultaneously to the touch-screen. Each [A] or [B] were the same icons as used for each subject in their previous phase, only increasing in number. Individual icons of the same category could not be distinguished one from another apart from their spatial location on the screen during any particular trial. The items of the array were to be interrogated and touched sequentially - all of category [A] first, followed by all of category [B]. The [AB] condition served as a control condition for the purpose of interpreting persistent error with the remaining three test conditions. Two, three and four error-types were possible for the test conditions [AAB], [ABB] and [AABB] respectively, and were interspersed with control [AB] trials at a ratio of one to every four test trials. The phase criterion was reached when fifteen out of the last (cumulative) twenty trials (75%) were correctly completed for each test condition, after which time that condition was no longer presented in any session of the phase. A session would be automatically halted when all of the four condition's criteria had been met.

Results. All six monkeys completed Phase 2. For each condition, the level of performance as set by the experimental criteria was reached, whilst maintaining 75% correct trials with the control condition. The number of trials, errors and percentage errors to criteria for the test conditions are shown in Tables 4.4-6.

Table 4.4. Nos. of trials, errors and % errors to criterion [AAB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	118	51	43.2%
Alfie	137	91	76.4%
Kissy	64	26	40.6%
Mimi	369	203	55.0%
Luba	102	57	55.9%
Ollie	113	51	45.1%

Table 4.5. Nos. of trials, errors and % errors to criterion [ABB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	38	13	34.2%
Alfie	24	8	33.3%
Kissy	30	8	26.7%
Mimi	24	9	47.5%
Luba	21	6	28.6%
Ollie	93	34	45.6%

Table 4.6. Nos. of trials, errors and % errors to criterion [AABB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	238	121	50.8%
Alfie	139	116	83.5%
Kissy	161	81	50.3%
Mimi	314	228	72.6%
Luba	300	174	58.0%
Ollie	376	217	57.7%

Discussion. The criteria performance measures have shown that all of the monkeys could successfully report three and four-item strings, independent of compositionality, using the same two category markers as presented in the previous training phase. It can also be seen that there is an order of completion effect across subjects. All completed the [ABB] condition first, the low numbers of errors demonstrating clear maintenance of the previous [AB] condition following the addition of a further [B] exemplar to be reported at the end of the new three-item set. Having maintained the serial order positions of the previous two-item set categories in the [ABB] condition, in comparison, the other two conditions [AAB and AABB] both insert another exemplar of [A] in the second position removing the option to succeed in the task by means of mere categorical chaining alone. Having now seen a clear demonstration by each monkey of a four-item sequence, it was now possible to expect that with training, the animal might attempt to order a three-item set composed of one novel and two familiar icons (i.e., that three items *per se* would not present a formidable problem).

Phase 3

Rationale. Having satisfactorily completed at least two three-item

sequences in the last phase, it was now not unreasonable to expect the monkey to attempt a further three-item string. With a view to the establishment of the core sequence for subsequent generalisation and multiplexing as shown above in figure 2.2, the three-item, three category string [ABC] was now presented for discrimination and explicit ordering. This second foundation phase would further inform us as to whether each monkey would be able to differentiate between its three particular stimuli in a reliable and consistently ordered manner.

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. Extending the sequence so as to provide a third class for subsequent expansion, this single training phase condition [ABC] employed an additional category [C] added to the existing two icon set [AB] seen in Phase 2, again discriminable by both colour and shape. An example screen array for this phase is shown in figure B.1. Each subjects' task was to sequentially interrogate and touch its category icons [A] first, followed by [B], then finally [C], with four error-types now possible (touching icon [B] or [C] first, [A] followed by [C] - a forwards error, or [A] [B] [A] - a reiteration). For this reason the experimental criteria for success with this 'core sequence' phase was very strict, again at the 75% level, but requiring a very high proficiency maintenance of ninety correct out of the last one hundred and twenty trials.

Results. All six monkeys completed this 'training' condition despite the higher level of performance required, and thus showed themselves capable of reporting a three-item, three category sequence. The number of trials, errors and percentage errors to criteria for the condition is shown in Table 4.7.

Table 4.7. Nos. of trials, errors and % errors to criterion [ABC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	731	397	54.3%
Alfie	199	89	44.7%
Kissy	356	188	52.8%
Mimi	1129	550	48.7%
Luba	300	109	36.3%
Ollie	195	75	38.5%

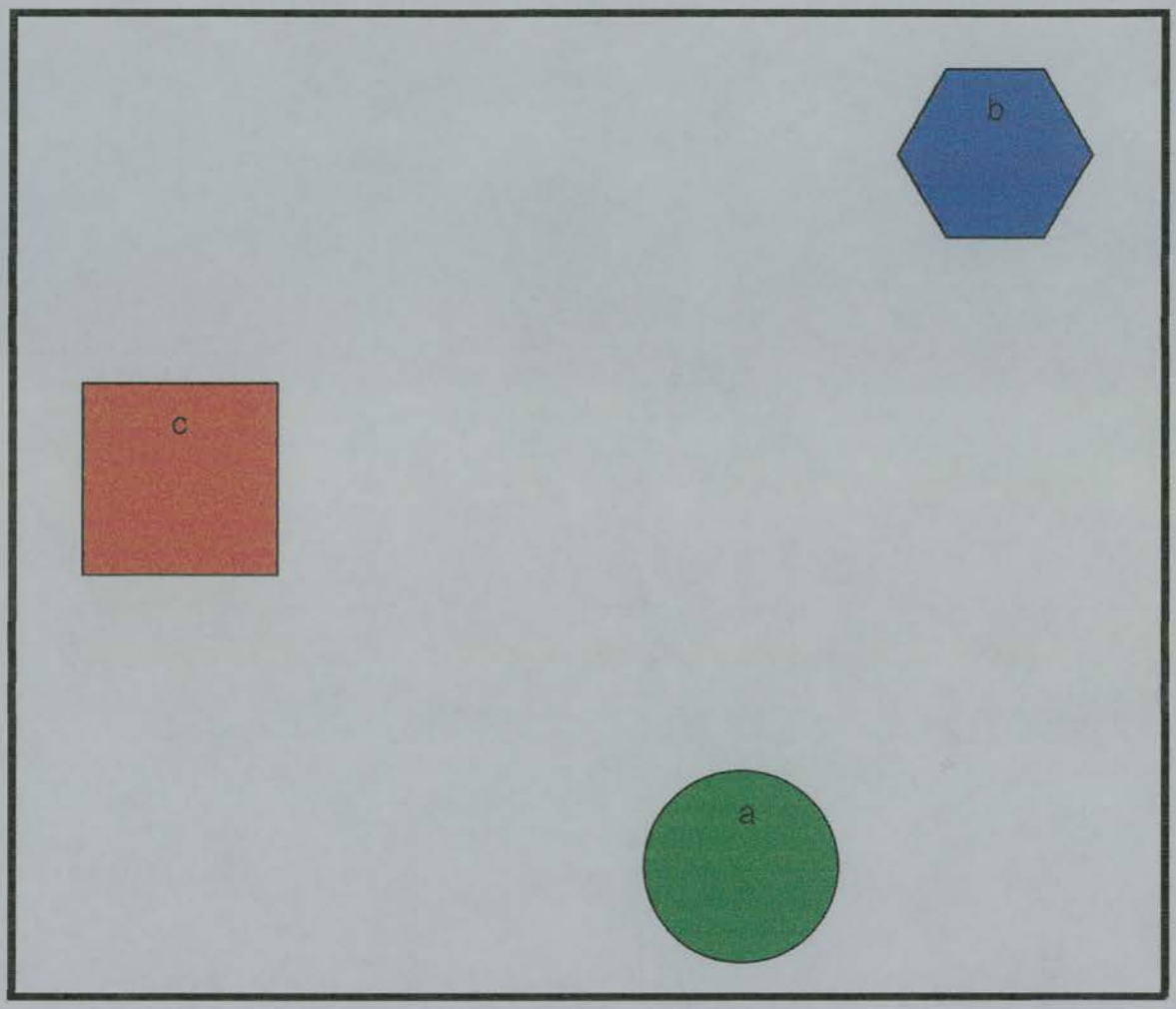


Figure B.1 Example of screen array for three-item set: [ABC]-core sequence (Phase 3)

Reaction time (RT) profiles for the last twenty correct trials (those for which the most consistent and reliable data may be obtained) indicated decreasing time to be taken by all of the monkeys to touch subsequent icons once they had begun to interrogate the touch-screen array. Between-subject variation was very low ($F_{5,120}$ = .47) with pooled RT mean data for the six animals showing significantly decreased times ($F_{2,120}$ = 5.59, p<.005) to touch subsequent icons as the sequence progressed (see Fig 4.1 below). A significant difference was found between the first and second (p=.05) but not the second and third category pooled mean values.

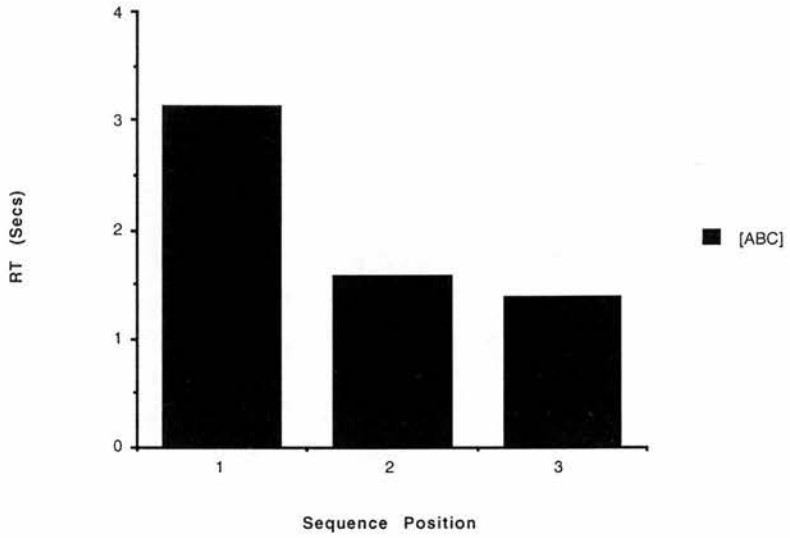


Figure 4.1 Mean RT profile for successful 3-item, three category [ABC] sequencing by Cebus monkeys (N=6, pooled)

Discussion. It has now been established that each monkey could consistently report their allocated 'core' three-item [ABC] sequence following a supervised training procedure which allowed no flexibility of ordering on the part of the subject. Compared to the earlier [AB] training phase, it can be seen that the extra loading of a third item has increased the number of trials required to reach the criterion level of performance, but in the presence of the enlarged error space (from one to four error types) this is perhaps not too surprising. For the pooled data, most of the errors recorded were of one of two types: either [AC] (38%) or [B] 1st (32 %), between them accounting for 70% of all the errors made in this [ABC] training phase. The low incidence of either the other two error typologies possible ([C] 1st or reiterative [ABA]) suggests that the monkey is

demonstrating some degree of conservation of class ordering from the earlier training phase [AB]. This idea may purchase some support from the RT findings of all six monkeys, in which the time taken to search each subsequent icon of the set to be ordered decreased throughout the sequence execution, suggestive of the development of a route-planning strategy to be taking place.

Phase 4

Rationale. With a view to determining whether the monkeys were going to be capable of the spontaneous classification of known familiar icons in an array, the next phase set out to explore the possibility of serial order information transfer to a novel six-item set, by increasing the sequence length (again by breadth) without changing the previously learned ordinal positions of the individual categories to be reported. With the addition of one item to each category simultaneously to produce a duplet six-item string [ABC] -> [AABBCC], whatever the outcome with regards to the spontaneous classification tests, savings measures could be computed by using a condition which requires the subject to first respond exhaustively to all items of the first category [A] before responding to all items of the second category [B] before all those of the remaining category [C], without the need for explicit within-category orderings. Doubling of the set size in this phase of the core-sequence expansion would not only offer an indication of the monkey's ability to spontaneously classify a principally orderable set, but would also, if successful, provide us with the first examples of a non-human primate proving itself to be capable of working with an ordered series (whatever the compositionality) of this size. Should the monkey succeed, however, such success would provide good evidence for the imposition of a classificatory scheme by the subject in order to assist the serial production of a multiple-item three-class series.

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. Having successfully completed the [ABC] training phases, all subjects qualified for this, the first of the probe transfer phases to be given (without a completion criterion). Introduced as short term probe trials, this phase was only offered the subject for a few

sessions, so preventing a loss of motivation for cases in which there might be too rapid an increased cognitive loading in the absence of sufficient task-success and reinforcement. The two conditions were presented at a ratio of 1:4 for single [ABC] and duplet [AABBCC] sequences respectively, the latter offering an error-typology of twelve possible paths with the six-item set. An example screen array for the six-item condition is shown in figure 4B.2. Following every twenty duplet string exposures, maintenance levels of three out of four successful [ABC] trials were required in order that the [AABBCC] conditions continue to be presented.

Results. For this transfer phase, although the subjects were not required to reached a criterion level of performance with the six-item set, all of the monkeys maintained the prerequisite 75% correct trial criterion for the control condition. The numbers of trials, errors and percentage errors for each condition were as shown in Tables 4.8-9.

Table 4.8. Nos. of trials, errors and % errors [ABC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	44	7	15.9%
Alfie	41	2	4.9%
Kissy	103	39	37.9%
Mimi	25	4	16.0%
Luba	37	4	10.8%
Ollie	35	4	11.4%

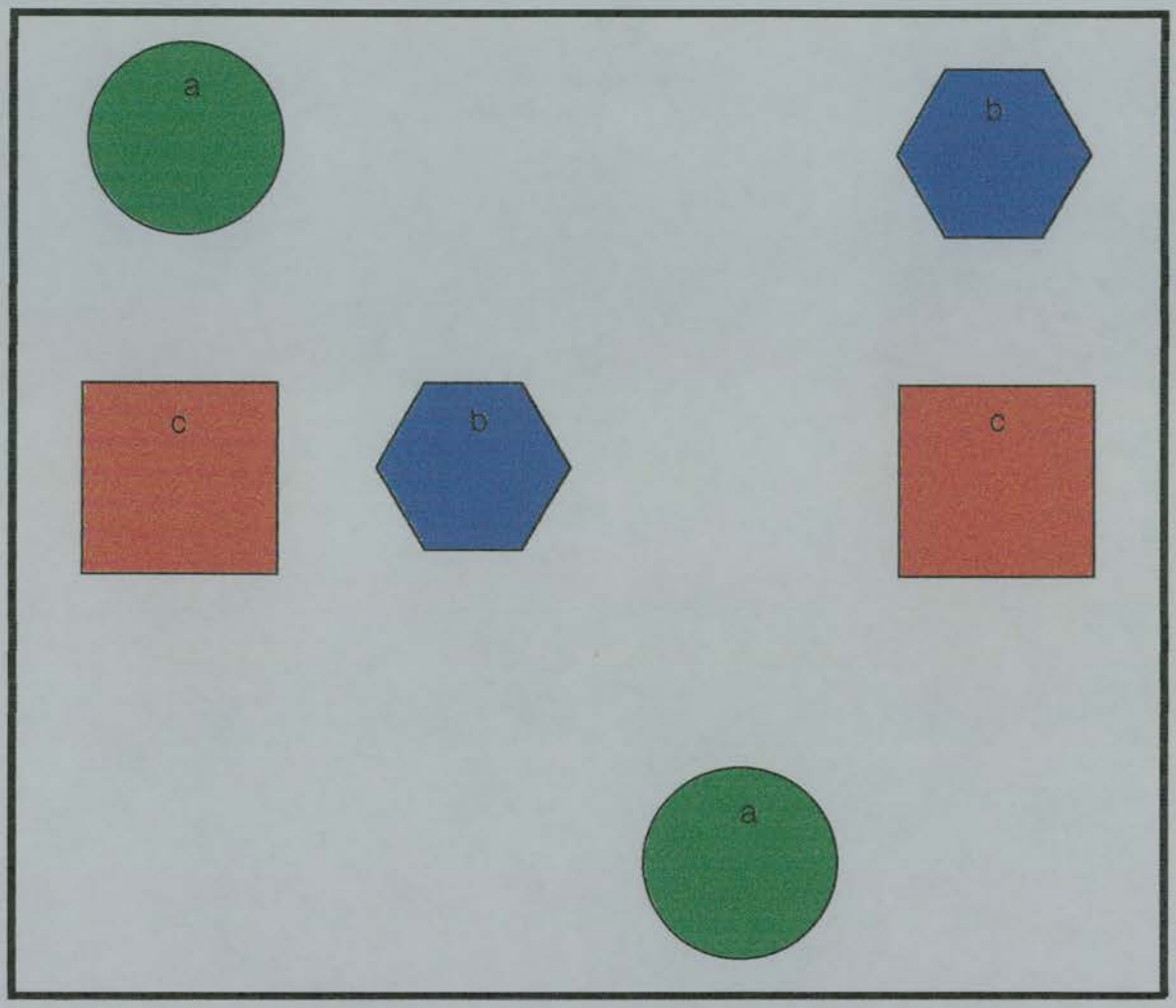


Figure B.2 Example of screen array for six-item set:
[AABBCC]-core sequence (Phase 4)

Table 4.9. Nos. of trials, errors and % errors [AABBCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	155	149	96.1%
Alfie	176	158	89.8%
Kissy	206	180	96.2%
Mimi	105	101	96.2%
Luba	81	62	76.5%
Ollie	135	116	85.9%

Of the possible error spaces for these two conditions, a first-touch error to either [B] or [C] was very rare, and although all error-types were represented in the 1531 incorrect trials' combined error variance, forward error now accounted for some 90% of the incorrect trials and of those, some 681 and 348 trials were of [AB]- and [AABC]-type incorrectly touched sequences respectively.

Discussion. Given that sequences of more than five items have hitherto not been reported in the literature for *any* non-human primate, it was not entirely clear whether such a serial production (whatever the compositionality of the set) might have been at all reportable by a monkey in this phase. The level of success achieved here with the new six-item sets (albeit low) clearly demonstrated the degree to which the monkey would indeed use a classificatory scheme to control and serially organise a sequence of multiple exemplars divisible into three distinct geometrical categories. Further, compared to its initial phase at acquisition (Phase 3), the [ABC] condition performance is almost now devoid of error for some subjects, showing an overall reduction of 20-30% in the number of incorrect trials. The 'core' sequence is thus holding up well in the face of the presentation of more complex stimuli arrays within the same session. For the six-item condition, all monkeys showed some degree of transfer success in their very first session despite receiving no explicit training on the task and indeed, 4 subjects showed at least one correct trial within the first 20 trials presented. Given the huge space of possible pathways through the new six-item string (see figure 3.2 above), one might not expect to see *any* correct trials at all during these early stages by chance measures alone, but it is also of significance to note that of the 12 possible error-types more than 67% may be accounted for by just two. These were the [AB] and [AABC] error typologies, both of which provide good evidence for attempts at conservation and spontaneous classification respectively,

the latter being made despite the occurrence of an error later in the sequence. Whatever the claims here for principled error making, it is remarkable that all of the monkeys clearly demonstrated the ability to spontaneously manage a six-item string at all, in the absence of explicit tuition or any differential feedback concerning how to do so.

Phase 5

Rationale. Given the rather large percentage of error trials for the previous Phase's single triplet to duplet step and doubling of sequence length to six items, and having consolidated to a high level the performance on the 'core' [ABC] trained sequence, it was now feasible to expect the monkeys to cope reasonably well with a lower incrementally increasing number of categorical exemplars once more on a free choice basis with just two categories. But would the subject continue to spontaneously classify by choosing all of the physically identical items of a category before the next, according to the category's ordinal position, independent of the number of exemplars being presented ? The following Phases 5 and 6 set out to answer this question without explicitly tutoring the animal in a single unique solution for each condition. This is so designed as to avoid the production of stereotyped motor responses whilst affording the need for continued array interrogation of interspersed conditions from trial to trial.

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. Four conditions [ABC], [AABB], [BBCC] and [AACC] comprised Phase 5, all six monkeys being presented with test conditions comprising two exemplars each of two categories ([A] and [B], [B] and [C] or [A and [C]) shown simultaneously to the touch-screen. Each [A], [B] or [C] were the same physically equivalent icons as used for each subject in their previous [ABC] phase, distinguishable only by spatial location. The item arrays of each condition were to be interrogated and touched sequentially - all of category [A] or [B] first, followed by all of either category [B] or [C]. The [ABC] condition served as a control condition for the purpose of interpreting performance with the remaining three test conditions. Four error-types were possible for each of the test conditions, and were interspersed with control [ABC] trials at a ratio of 1:4 for each of

the test trials. The phase criterion was reached when fifteen out of the last (cumulative) twenty trials were correctly completed for each test condition, after which time that condition was no longer presented in any sessions of the phase. The session was automatically halted when all four condition criteria had been met.

Results. For each condition, all of the six monkeys reached the level of performance as set by the experimental criteria, whilst maintaining 75% correct trials with the control condition. The number of trials, errors and percentage errors to criteria for the test conditions of Phase 5 are shown in Tables 4.10-13.

Table 4.10. Nos. of trials, errors and % errors to criterion [ABC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	24	0	0.0%
Alfie	10	0	0.0%
Kissy	15	2	3.3%
Mimi	82	8	9.8%
Luba	29	8	27.6%
Ollie	8	0	0.0%

Table 4.11. Nos. of trials, errors and % errors to criterion [AABB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	108	55	50.9%
Alfie	42	13	31.0%
Kissy	77	31	40.3%
Mimi	290	186	64.1%
Luba	97	44	45.4%
Ollie	39	15	38.5%

Table 4.12. Nos. of trials, errors and % errors to criterion [BBCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	118	58	49.2%
Alfie	23	8	34.8%
Kissy	49	20	40.8%
Mimi	557	373	77.0%
Luba	102	58	56.9%
Ollie	54	26	48.1%

Tables 4.14-17 show a summary of Phase 5 error profiles for each monkey and actual touch sequences for incorrect trials. The frequency count for a total of four possible error typologies are given for each condition.

Table 4.13. Nos. of trials, errors and % errors to criterion [AACC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	36	13	36.1%
Alfie	52	22	42.3%
Kissy	20	5	25.0%
Mimi	202	110	54.5%
Luba	97	44	45.4%
Ollie	18	3	16.7%

Table 4.14. Individual monkey error profiles [ABC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	0	0	1	5	1	0
C 1st	0	0	0	0	1	0
AC	0	0	1	3	5	0
ABA	0	0	0	0	1	0

Table 4.15. Individual monkey error profiles [AABB]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	1	0	5	34	11	0
AB	53	12	24	147	30	15
AAA	1	1	0	1	0	0
AABA	0	0	2	4	3	0

Table 4.16. Individual monkey error profiles [BBCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
C 1st	2	0	1	54	6	1
BC	24	5	12	183	33	15
BBB	16	0	0	39	4	2
BBCB	16	3	7	97	15	8

Table 4.17. Individual monkey error profiles [AACC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
C 1st	0	2	0	12	0	0
AC	6	6	1	85	16	2
AAA	6	12	2	9	3	0
AACA	1	2	2	4	10	1

Discussion. The criteria performance measures have shown that all of the monkeys will successfully report four-item duplet strings, independent of compositionality, using two of the three category markers with each maintaining their relative sequence order as presented in the previous [ABC] training phase (although possibly at variance with their absolute serial positions). It can also be seen that there was an order of completion effect for each subject. All completed the [AACC] condition first, with the fewest number of trials and percentage errors to criterion, suggestive of a privileged cognisance of the 'end-anchoring' positions of [A] and [C] within the [ABC] learned core-sequence. Likewise, first touch errors were recorded on average only 10% of the time, especially salient now that the insertion of a second exemplar for each category in the second position had also removed the option to succeed in the task by means of categorical chaining alone. Despite the possibility of four error types for each of the test conditions, error profiles however reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 50% of all error trials for the remaining conditions for which this strategy did not allow serial success. Whether the monkey was attempting to exhaustively search the array by reporting, say, [ABAB] for the [AABB] condition, cannot be known at this stage due to the constraints being imposed upon the subject by the tasks' programming requirements.

Phase 6

Rationale. Having now seen a clear demonstration by each monkey of consistent performance with two-category four-item sequences, it was now possible to assess whether, without changing the icon set, the subject might attempt to order the existing sequences extended in breadth once more. It had already seen in Phase 4 that all of the monkeys were capable of achieving some success with a sequence length of six items, the minimum string length required for the meaningful introduction of any depth of search for the later hierarchical phases of the experimental series as previously discussed. Given the monkeys' successes in the last phase, it seemed that the next step would be to increase the number of categorical exemplars once more, still on a free choice basis, using the same two-category string sequences. An additional condition also presented a six-item sequence, but this time with a duplet composition

using all three categories. This interspersion of duplet and triplet six-item sequences offered the possibility of assessing the subject's performance with equal string-lengths differing in their category's serial and ordinal positions. The aim was to assess the degree to which the subject might continue to classify by choosing all of the physically identical items of a category before the next, according to the *category's* fixed ordinal positions, again independent of the number of equivalent exemplars being presented.

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. Phase 6 contained five interspersed conditions, [ABC], [AAABBB], [BBBCCC], [AAACCC] and [AABBCC] using one, two or three exemplars of the two or three categories shown simultaneously to the touch-screen as with the previous phase. Each [A], [B] or [C] were the same icons as used for each subject in their core [ABC] sequence, only increasing in number within any given string. Individual icons of the same category could not be distinguished one from another apart from their spatial location on the screen. The items of the array were to be interrogated and touched sequentially - all of category [A] or [B] first, followed by all of category [B] and then all those of category [C]. The [ABC] condition again served as a control condition for the purpose of evaluating persistent error within the remaining four test conditions. As many as eight or twelve error-types were now possible for the triplet and duplet test conditions respectively, and were interspersed with control [ABC] trials at a ratio of 1:4 of each of the five test trials. The phase criterion was reached when fifteen out of the last (cumulative) twenty trials were correctly completed for each test condition, after which time that condition was no longer presented in any session of the phase. The sessions were automatically halted when the criteria for all five conditions had been reached.

Results. For each condition, all six monkeys succeeded at reaching the level of performance as set by the experimental criteria, whilst maintaining 75% correct trials with the control condition. The mean number of trials, errors and percentage errors to criteria for the test conditions are shown in Tables 4.18 - 21.

Table 4.18. Nos. of trials, errors and % errors to criterion [AAABBB]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	261	211	80.8%
Alfie	214	147	68.7%
Kissy	455	300	67.4%
Mimi	250	187	74.8%
Luba	182	95	52.2%
Ollie	664	447	67.3%

Table 4.19. Nos. of trials, errors and % errors to criterion [BBBCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	163	115	70.5%
Alfie	449	290	64.6%
Kissy	272	161	59.2%
Mimi	1525	1053	55.5%
Luba	60	33	55.0%
Ollie	394	242	61.4%

Table 4.20. Nos. of trials, errors and % errors to criterion [AAACCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	209	167	79.9%
Alfie	121	63	52.0%
Kissy	218	126	57.8%
Mimi	195	114	58.5%
Luba	109	54	49.5%
Ollie	20	5	25.0%

Table 4.21. Nos. of trials, errors and % errors to criterion [AABBCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	377	288	76.4%
Alfie	537	347	64.6%
Kissy	523	342	65.4%
Mimi	428	301	70.3%
Luba	272	197	72.4%
Ollie	851	593	39.7%

Four of the six monkeys completed the [AAACCC] condition first, the remaining two (Charlie and Luba) first completing [BBBCCC]. All except one (Mimi) ended the phase with the duplet string [AABBCC] although with an equivalent percentage error rate. Tables 4.22-25 show a summary of Phase 6 error profiles for each monkey and actual touch sequences for incorrect trials. The frequency count for a total of eight or twelve possible error typologies are given for triplet and duplet conditions respectively.

Table 4.22. Individual monkey error profiles [AAABBB]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	43	3	24	9	1	24
AB	57	31	108	29	14	110
AAB	86	84	145	93	43	290
AAA*	9	10	5	18	7	10
AAAA*	5	8	5	11	19	10
AAABA	0	2	0	1	3	1
AAABBA	0	1	10	2	0	0
AAABBB*	11	7	3	24	8	2

* = Reiteration error

Table 4.23. Individual monkey error profiles [BBBCCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
C 1st	43	3	24	9	1	24
BC	57	31	108	29	14	110
BBC	86	84	145	93	43	290
BBB*	9	10	5	18	7	10
BBBB*	5	8	5	11	19	10
BBBCB	0	2	0	1	3	1
BBBCCB	0	1	10	2	0	0
BBBCCC*	11	7	3	24	8	2

* = Reiteration error

Table 4.24. Individual monkey error profiles [AAACCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
C 1st	21	1	4	1	2	0
AC	48	5	11	6	3	3
AAC	69	30	75	53	32	3
AAA*	10	4	3	22	2	1
AAAA*	8	12	10	11	5	1
AAACA	0	0	2	1	0	0
AAACCA	0	0	7	2	0	0
AAACCC*	11	11	8	17	10	0

* = Reiteration error

For the triplet string conditions of the phases [AAABBB], [BBBCCC] and [AAACCC], there was always a single error type which could account for around 50% of each condition's incorrect trials. These dominant error types were of the same relative composition for each condition, reflecting

a forwards-error of one item across the category boundary. For example, the dominant error typology for condition [AAABBB] was to produce a sequence touch of [AAB], for [BBBCCC] it was [BBC]. However, for the three-duplet condition [AABBCC] a different profile was to emerge (see Table 4.25). Although there was now twelve possible error types, more than 63% of all the incorrect trials were accounted for by only two error sequences. As with the triplet conditions, typical of all the monkey's performances, pooled errors for the [AABBCC] interspersed duplet condition occurred at category boundaries, mostly as a result of forwards-error touching response sequences [AB] (36%) or [AABC] (27%).

Table 4.25. Individual monkey error profiles [AABBCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	56	3	56	9	7	60
C 1st	13	2	2	2	2	9
AB	132	104	172	77	49	272
AC	6	8	6	4	5	14
AAA*	10	35	13	34	23	21
AAC	7	19	4	10	25	17
AABA	1	11	5	3	2	5
AABC	39	132	62	81	78	127
AABBA	0	2	6	1	1	7
AABBB*	8	13	6	25	5	11
AABBCA	1	2	5	3	0	2
AABBCB	15	15	1	52	3	48

* = Reiteration error

A further difference between the duplet and triplet six-item sequences was revealed by examination of the RT profiles for correct trials. As may be seen in figures 4.2-3, using data from the same monkey (Luba), another indication of the significance of category boundaries was becoming apparent. Whereas the RT measures of correctly executed triplet [BBBCCC] showed a significantly increased search time ($p<.05$) only for the fourth position, the duplet [AABBCC] condition resulted in significant search time increases at the third and fifth ordinal positions.

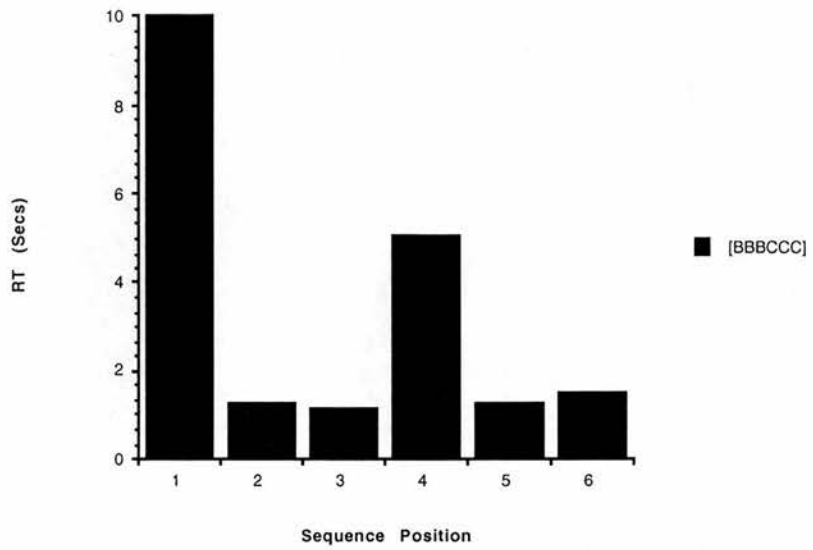


Figure 4.2 Mean RT profile for successful 6-item sequencing of two categories, each with three exemplars by a Cebus monkey (Luba)

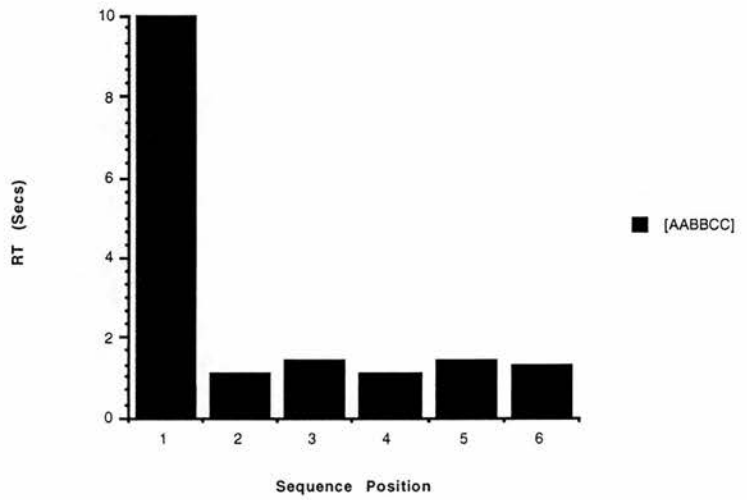


Figure 4.3 Mean RT profile for successful 6-item sequencing of three categories, each with two exemplars [AABBCC] by a Cebus monkey (Luba)

Discussion. Condition criteria were met by all of the monkeys in this phase, successfully reporting any one of four interspersed six-item triplet and duplet strings. This success was readily upheld independent of compositionality, and using two or three of the category markers with each maintaining their relative sequence order as presented in the previous [ABC] training phase (although again possibly at variance with

their absolute ordinal positions). It can also be seen that there was once more an order of completion effect for each subject. Most of the monkeys completed the [AAACCC] condition first, with the fewest number of trials and percentage errors to criterion, further supporting the idea of there being a privileged cognisance of the 'end-anchoring' positions of [A] and [C] within the [ABC] learned core-sequence. Likewise, first touch errors were recorded on average only 8% of the time, especially notable now that the insertion of a third exemplar for each category in the third position had further removed the option of succeeding in the task by means of categorical chaining alone. Despite the possibility of eight or even twelve error typologies for each of the test conditions, error profiles revealed clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing once more, their accounting for more than 63% of all the error trials for combined triplet conditions. Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner cannot be known due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (any forwards-error touch ended the trial). However, it was now becoming quite clear that the *Cebus* monkeys were quite capable of demonstrating their proficiency with six-item sequences under supervised-learning conditions for which a degree of free search was permissible. Evidence for the existence and use of some strategic planning (or at least economic management) with these larger sequences come not only from an analysis of errors frequency and the shift towards dominantly forward looking errors, but were also becoming evident following the development of clearly differentiable RT profiles for six-item sets with variant compositionality. If the RT profiles were to reflect merely sequence position *per se* for the first and subsequent touches up until the end of the string, one would not expect to record any differences relating to the composition of a sequence of identical length. However, the findings indicate not only that compositionality is indeed a factor in determining the nature of the RT profile for sequence positions, but that the profile contour produced appears to reflect the nature of categorical boundary features of the condition being interrogated. Further, much of the time taken to successfully complete each trial was used up in the period before the first touch had been made, indicative of careful inspection and forward planning prior to sequence execution. These findings possibly provide the first indications of a further index of

classificatory processes to be revealed in the monkey. If this were indeed to be true, then such phrasing effects might be expected to persist (or to become even more salient) with increasingly larger string-length and its compositional variations.

Phase 7

Rationale. Having by now satisfactorily completed four six-item sequences (whether composed of two or three categories) in the last phase, it was now possible to conceive that the monkey might readily attempt strings with an additional third exemplar to each category, now containing three identical items. Introduced as short term probe trials, this phase was only offered the subject for a few sessions, so preventing too large an increase in cognitive loading without sufficient task-success and reinforcement. This second generalisation/transfer phase would further inform us as to whether any of the monkeys would be able to readily differentiate between the components of such a large stimulus array, and indeed be capable of executing a nine-item sequence at all, let alone to do so in any systematically controlled or principled way.

Subjects. All six adult monkeys took part in this phase, despite one of the females (Mimi) giving birth to a second infant in her colony room, her first settling to his rehousing in an adjacent open-colony 'bachelor' group. Much colony calling was taking place within the laboratory environment at this time of the transitions, but disruption to the monkey's performance in the experimental test-rooms was short-lived.

Design and Procedure. Extending each monkey's core [ABC] sequence further again, this single condition probe phase introduced the monkey to a nine-item set for the first time, cold. Adding three exemplars of the third category [C] to the previously successful two category triplet icon set [AAABBB] seen in Phase 6, the new sequence was composed of three categories, each with three identical exemplars [AAABBBCCC]. Each subjects' task was to sequentially interrogate and touch all of its category [A] icons first, followed by all of the [B], then finally all of the [C]. There was no experimental criterion for success with this probe phase which would be presented for only two or three consecutive sessions according to the subjects disposition to the task, each presenting a maximum of fifty

trials per session.

Results. All six monkeys completed at least one session and succeeded in reporting correct sequencing of this nine-item condition despite the much higher level of performance required. The number of trials, errors and percentage errors (without phase criteria) for this condition is shown in Table 4.26-27.

Table 4.26. Total numbers of probe trials, errors and % errors [AAABBBCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	59	54	91.5%
Alfie	104	94	90.4%
Kissy	79	66	83.5%
Mimi	75	72	96.0%
Luba	163	144	88.3%
Ollie	111	84	75.7%

Table 4.27. Individual monkey error profiles [AAABBBCCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	3	1	1	0	0	1
C 1st	2	1	0	1	4	0
AB	6	9	9	7	8	4
AC	1	1	0	1	0	0
AAA*	2	0	2	0	2	2
AAB	19	54	31	29	52	37
AAC	2	0	0	3	1	0
AAAA*	2	1	1	3	12	3
AAAC	2	1	0	2	1	0
AAABA	0	0	0	0	3	1
AAABC	3	11	1	13	9	2
AAABBA	0	0	1	1	1	0
AAABBB*	1	1	0	0	2	2
AAABBC	7	11	13	12	34	24
AAABBBA	0	0	1	0	1	0
AAABBBB*	1	1	1	0	5	1
AAABBBCA	0	0	1	0	0	0
AAABBBCB	0	0	0	0	0	3
AAABBBCCA	0	0	0	0	0	0
AAABBBCCB	0	0	0	0	0	2
AAABBBCCC*	3	3	3	0	4	2

* = Reiteration error

Discussion. Although there were no experimental condition criteria for this phase, every monkey succeeded in reporting correct sequencing of a nine-item string despite the much higher level of difficulty and performance required. This has provided the first demonstration of the abilities of a non-human primate to include the reporting of any sequence of this length, let alone a principled successful execution of a simultaneous three category, three-exemplar sequence, as was presented in the stimuli array for this condition. Despite the relatively low number of trials presented to each monkey in this phase, an indication of their successfully exploiting the principle of the task's simple classificatory scheme was to be seen both in the fact that all succeeded in completing a correct sequence within the first twenty trials, and that there were significant patterns in their error profile for incorrect trials. First touch errors, which with a nine-item array one might expect to see reported more frequently now, were recorded for less than 1% of the total error space for all subjects. Given the possibility of as many as twenty-one error typologies for this new nine-item test condition, error profiles revealed clear preferences for sequential touching in accordance with target category sequencing once more, with one-step forwards-errors [AAB] and [AAABBC] accounting for more than 63% of all the error trials recorded (at a rate of 42% and 21% respectively). Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner still remains unknown at this stage due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (any forwards-error touch ended the trial). Evidence for the existence and use of some strategic planning (or at least economic management) with movement though this much larger sequence comes only from an analysis of error frequency. For the very few correct trials, much of the time taken to successfully complete each trial continued to be represented in the initial reaction time to the first touch (IRT), indicative of careful inspection and forward planning prior to sequence execution, but otherwise did not reveal any consistent pattern of significance beyond the fourth serial position. Nonetheless, these findings provide the first indications of the successful execution of a nine-item sequence by a non-human primate (or any other species for that matter) and indicate a level of performance consistent with the presence of classificatory processes in the monkey, at least in a supervised laboratory environment. If this were indeed to be true, then such

phrasing effects might be expected to persist (or to become even more salient) with increasingly larger string-length and its compositional variations.

Phase 8

Rationale. Having by now satisfactorily demonstrated the ability to negotiate a nine-item sequence composed of three categories, each with three exemplars in the last phase, it was now possible to give the monkey another training phase with a view to establishing the development of the core [ABC] sequence so as to produce a string of sufficient length suitable for the later introduction of a depth of search, within a category (see figure 3.3). This fourth training phase (the previous three being [AB] , [ABC] and [AABBCC]) would further inform us as to whether any of the monkeys would be able to consistently report a novel nine-item stimulus array in a systematically controlled and principled way.

Subjects. All six adult monkeys took part in this phase

Design and Procedure. Continuing with the extended [AAABBBCCC] sequence for each monkey, this phase presented interspersed trials of control [AABBCC] to assist interpretation of any failure to maintain or improve performance with the nine-item sequence. As in the previous phase, each subjects' task was to sequentially interrogate and touch all of its category [A] icons first, followed by all of the [B], then finally all of the third category [C]. The two conditions were presented at a ratio of 1:4 for duplet and triplet sequences respectively, employing multiple exemplars of the same icons of categories [A], [B] and [C] as previously used. An example array for the nine-item condition is shown in figure 4B.3. Following every twenty triplet string exposures, maintenance levels of three out of four successful [AABBCC] trials were also required in order that the [AAABBBCCC] condition continue to be presented. The phase was completed when the triplet condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

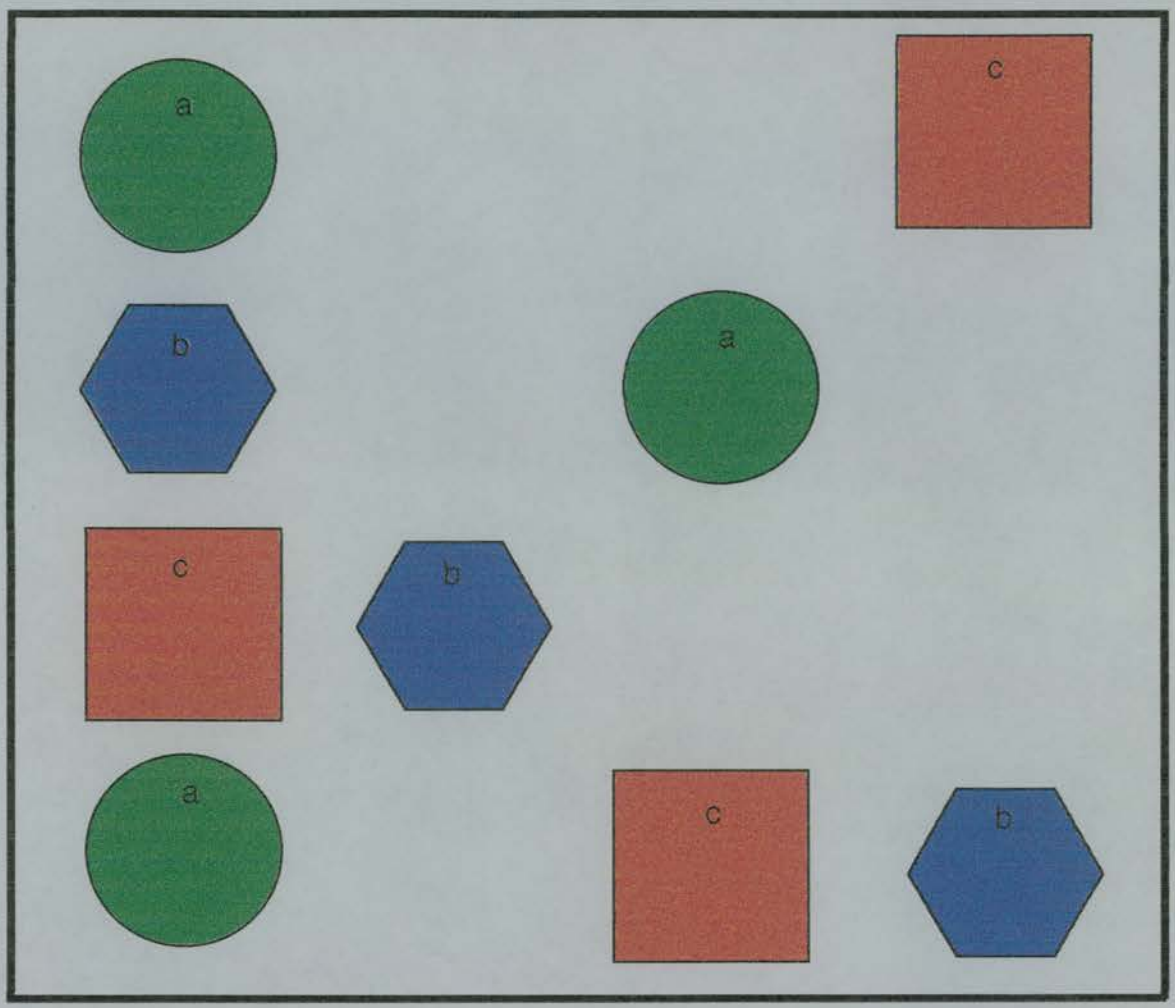


Figure B.3 Example of screen array for nine-item set: [AAABBBCCC] (Phase 8)

Results. All of the monkeys completed this 'training' condition and were thus eligible for continuation with the experimental series. The 75% criterion required for the control condition [AABCC] exposures were consistently upheld by all subjects for the duration of the phase and the mean number of trials, errors and percentage errors to criteria for the test condition were as shown in Table 4.28.

Table 4.28. Nos. of trials, errors and % errors to criterion [AAABBBCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	1686	1114	66.0%
Alfie	2801	1993	71.1%
Kissy	2514	1956	77.8%
Mimi	1950	1659	85.0%
Luba	1339	1120	83.6%
Ollie	1382	841	60.8%

Still accounting for between 71% and 85% of the total trials for any one subject, it may be seen from Table 4.29 that although more of the error typologies are represented were compared with the previous phase, more than 66% of the total number of error trials could nonetheless be accounted for by only two of the possible twenty-one error types. (The error profile for the control condition [AABCC] were roughly comparable with those reported for Phase 6, with the two dominant error types maintaining their share of the total error space at a level of 62% for typologies [AB] - 32% and [AABC] - 30% respectively.)

Table 4.29. Individual monkey error profiles [AAABBBCCC]

Error types	Charlie	Alfie	Kissy	Mimi	Luba	Ollie
B 1st	5	39	144	94	15	1
C 1st	11	17	26	20	2	0
AB	47	106	117	112	89	47
AC	6	12	10	24	8	1
AAA*	38	45	54	51	36	26
AAB	400	771	734	562	462	365
AAC	8	13	19	25	9	0
AAAA*	58	68	58	79	26	55
AAAC	12	20	24	21	14	0
AAABA	14	11	31	16	7	5
AAABC	39	85	105	102	72	15
AAABBA	10	8	27	18	10	2
AAABBB*	23	25	16	24	11	28
AAABBC	311	549	461	314	238	245

AAABBBA	13	7	7	13	4	3
AAABBBB*	37	48	23	63	24	4
AAABBBCA	7	5	3	4	4	0
AAABBBCB	4	9	9	18	1	4
AAABBBCCA	5	5	2	8	3	1
AAABBBCCB	5	7	5	19	3	4
AAABBBCCC*	0	0	0	0	0	0

* = Reiteration error

With insignificant variation between monkeys ($F_{5,48}$ = .65) the mean reaction times profile for pooled data for the nine-item condition was as shown in figure 4.4. Significant increases in RT were never seen to occur between the intra-category values recorded, but only those occurring at the fourth and seventh sequence positions (both lying at the category boundaries) showed significantly increasing times to icon touching ($p<.05$ and $p<.001$ respectively).

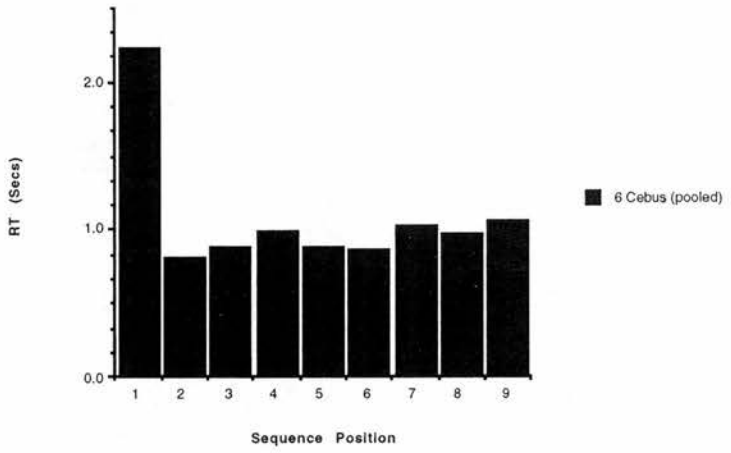


Figure 4.4 Mean RT profile for successful 9-item sequencing of three categories, each with three exemplars [AAABBBCCC] by Cebus monkeys (N = 6, pooled)

Discarding the IRT, such analysis thus indicates significant differences to occur only at both the AB and BC category boundaries.

Discussion. Condition criteria were met by all of the monkeys in this phase, successfully reporting both six- and nine item duplet and triplet sequences. This success was consistently demonstrated independent of compositionality, using three category markers with each maintaining their relative sequence order as presented in their previous [ABC]

training phase although with different absolute serial positions. First touch errors were recorded less than 4% of the time, indicating a high degree of inspection of the now greatly expanded array despite the greatly decreased initial reaction time. Despite the possibility of twenty-one error typologies for the test condition, error profiles revealed clear preferences for touching in accordance with the previous [AB] and [BC] category sequencing once more, with two forward-error types accounting for more than 63% of all the error trials for the three-triplet condition. Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner remains unknown due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (a forwards-error touch ending the trial). Further evidence for the existence and use of some strategic planning (or at least economic management) with this larger sequence comes not only from an analysis of errors frequency, but is supported by the finding of clearly differentiable RT profiles for the nine-item set. The findings indicate not only that compositionality is indeed a factor in determining the nature of the RT profile for sequence positions, but that the profile contour produced appears to reflects the nature of categorical boundary features of the condition being interrogated. Such phrasing effects, in which the search of items within a class is generally much faster than for that of the first item of a class, is akin to a form of chunking, indicative of the ongoing development of classificatory expertise in the monkey.

Interim Summary (Phases 1-8)

It has now become quite clear that the *apella* monkey is quite capable of demonstrating the execution of an exhaustive search procedure with sequences up to, and including, nine items under supervised-learning conditions for which a degree of free search was permissible. Six monkeys have now been taken through an incremental series of steps comprising some eight phases using a core, three-item [ABC] icon set subsequently expanded in breadth to provide the nine-item, three-icon, three-exemplar multiplexed set of Phase 8. Well beyond the previously recorded sequence lengths for any serial-order production success by a non-human primate, none of the individual subjects participating in the experimental series to date have failed to demonstrate success with any of the phases so far

presented. This includes the attainment of the criterion level of performance set for the large nine-item, three category sets, and current performances offer no indication that this might be anywhere close to the upper limit of serial-order control execution by these subjects. Furthermore, given that each increase in sequence length simultaneously increases the difficulty as indexed by the combinatorial expansions in the number of possible pathways (see fig. 3.2, and thus the number of errors) through the set, and, with the condition criteria remaining at the seventy-five percent level, all of the monkeys are actually demonstrating accelerating levels of performance in the absence of the proportional increase in the numbers of errors to be expected. Indeed as can be seen from the summary tables 4.30-32 below, all subjects succeeded at all task levels without fail, actually improving their relative performances as measured by error trials to criterion for the increasingly more difficult sequences (3-, 6- and 9-item set data show considerably less frequent, even falling, errors than might be expected by chance projections alone).

Table 4.30. Nos. of trials and errors to criterion [ABC]

Subject	No. of Trials	No. of Errors	Errors Projected
Charlie	731	397	1680
Alfie	199	89	1680
Kissy	356	188	1680
Mimi	1129	550	1680
Luba	300	109	1680
Ollie	195	75	1680

Table 4.31. Nos. of trials and errors to criterion [AABBCC]

Subject	No. of Trials	No. of Errors	Errors Projected
Charlie	155	149	51200
Alfie	176	158	51200
Kissy	206	180	51200
Mimi	105	101	51200
Luba	81	62	51200
Ollie	135	116	51200

Table 4.32. Nos. of trials and errors to criterion [AAABBBCCC]

Subject	No. of Trials	No. of Errors	Errors Projected
Charlie	1686	1114	9640000
Alfie	2801	1993	9640000
Kissy	2514	1956	9640000
Mimi	1950	1659	9640000
Luba	1339	1120	9640000
Ollie	1382	841	9640000

Not only do they thus continue to succeed with these more difficult and complex tasks (never before attempted within a single paradigm), the monkeys are showing this success despite their having to perform at increasingly higher levels of competence in the face of what is actually a decreasing level of reward. (In real terms, the amount of effort required to achieve a seventy-five percent criterion with a nine-item set is very much more costly than that required to reach the same criterion level with a only three-item set. However, the reward, a single peanut, has remained the same throughout !).

Despite the early fears that the subjects motivation might thus become lessened as the stakes were raised, one further indication that the monkey was starting to tune into salient features of the multiplexed sequences was to be derived from an analysis of latency data, Not only were the monkeys sufficiently motivated and consistent in their engagement with the task throughout each individual trial (interrupted sequences were extremely rare), the reaction time profiles for each and every subject incidentally showed specific sequence position effects coincident with the category boundaries. Not specifically a variable controlled for in the task design, the RT profiles were not found to merely reflect sequence item position selection from the screen array, and one might not necessarily have expected to record any differences relating to the composition of the string under review. However, as was seen in the case of at least the six-item sets, a single phase may provide rather different profiles for sequences of the same length, yet differing in their compositionalities. Furthermore, not only did this finding (see figs 4.2-3) show that the RT profile not to be merely a function of sequence position *per se*, it was found that significant increases in reaction time were to occur only at the category boundaries. The evidence for this category boundary effect becomes most apparent when one realises that not only does the RT profile shift position consistent with the compositionality of sequences of otherwise equivalent string length, this category boundary effect is seen to reliably occur in the same subject, during the execution of these interspersed multi-category conditions within the same phase. In the latest of the phases to have been reported(the 9-item set of Phase 8) it was found that the RT profile again showed significant increases only at these categorical boundaries, indicative of the monkeys' relaxing into a

patterned 'chunking' response suggestive of the need to overcome the increasing memorial constraints otherwise imposed by the increasingly longer sequence lengths. Taken all together, these findings have provided the first convincing evidence for the control of spontaneous classificatory processes in the monkey. The following phases were designed to further investigate the nature and extent of this well adapated classification skill seen to be emerging in all of the laboratory monkeys so far tested.

Phase 9

Rationale. Having now seen clear demonstrations by each monkey of consistent nine-item sequencing using three-categories, each of three-exemplars, it was now possible to assess whether, again without changing the icon set, the subject might attempt to continue ordering the existing sequences when extended in breadth once more. Although it was now known that the monkeys were capable of achieving success with a sequence length satisfying the minimum string length required for the meaningful introduction of any depth of search, we do not know what the upper limit might be for the monkey's string length when extended in breadth alone. Given the monkeys' successes of the last phase, it seemed that a next exploratory step might be to increase the number of categorical exemplars once more, still on a free choice basis, using the same three-category string as a base-sequence. Incrementing the sequence length by the addition of one item, the aim was to assess the degree to which the subject might continue to demonstrate classification by choosing all of the physically identical items of a category before the next, according to the category's fixed serial positions, independent of its ordinal position and the number of equivalent exemplars contained within the sequence being presented.

Subjects. Of the six monkeys qualifying for Phase 9, only five went on to see a set of three interspersed ten-item strings with category exemplar numbers varying between three and four identical icons per class. [Luba, who had just given birth again, was not tested for some six weeks, after which time all testing had ceased for about eight months during a period of extensive laboratory refurbishment].

Design and Procedure. Phase 9 contained three interspersed conditions, [AAAABBBCCC, AAABBBBCCC & AAABBBCCCC] in each of which there were four exemplars of one category and three each of two other categories. Each [A], [B] or [C] were the same icons as used for each subject in their previous sequences, only increasing in number within any given string. Individual icons of the same category could not be distinguished one from another apart from their spatial location on the screen. The items of the array were to be interrogated and touched sequentially - all of category [A] or [B] first, followed by all of category [B] and then all those of category [C], with as many as twenty-four error-types now possible for the new test conditions. The phase criteria was reached when fifteen out of the last (cumulative) twenty trials were correctly completed for each test condition, after which time that condition was no longer presented in any session of the phase. The sessions were automatically halted when the criteria for all three conditions had been reached.

Results. For each of the three ten-item conditions in this phase, all five monkeys reached the levels of performance as set by the experimental criteria. The mean number of trials, errors and percentage errors to criteria for the test conditions are shown in the following Tables 4.33-35.

Table 4.33. Nos. of trials, errors and % errors to criterion [AAAABBBCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	212	110	51.9%
Alfie	948	594	62.6%
Kissy	402	246	61.2%
Mimi	511	393	76.9%
Ollie	453	316	69.7%

Table 4.34. Nos. of trials, errors and % errors to criterion [AAABBBBCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	714	434	60.8%
Alfie	1206	767	63.6%
Kissy	991	563	56.8%
Mimi	500	367	73.4%
Ollie	405	275	67.9%

Table 4.35 Nos. of trials, errors and % errors to criterion [AAABBBCCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	105	68	64.7%
Alfie	59	33	55.9%
Kissy	258	142	55.0%
Mimi	41	21	51.2%
Ollie	401	247	61.6%

All five monkeys completed the [AAABBBCCCC] condition first (showing the fewest number of trials and the lowest percentage of errors to criterion). Of the remaining two conditions, four monkeys completed the [AAAABBBCCC] second before completing the phase with a criterion run with [AAABBBBCCC] condition. Although a total of twenty-four error typologies were possible for each ten-item condition, there was actually a single error string which accounted for more than 30% of the incorrect trials. A second error typology could account for at least a further 20% or so. An exhaustive error profile is not provided here, but Table 4.36 shows the dominant error types for the three conditions.

Table 4.36. Dominant error types for three ten-item sequences

Condition	Dominant Error	Second Error	% of Total Errors
AAAABBBCCC	AAAB	AAAABBC	63.9%
AAABBBBCCC	AAABBBC	AAB	61.1%
AAABBBCCCC	AAABBC	AAB	55.5%

Reaction time data for correct trials reveals differences between the ten-item sequences, each varying according to compositionality. For any given sequence, pausing at the category boundary was common, as may be seen in figures 4.5-7, again using data from the same monkey (this time Ollie). RT measures of correctly executed triplet and quadruplet category exemplars within a sequence showed increased search times for the first item of each class to be touched, irrespective of ordinal position within the ten-item set being interrogated. As with the six-item sets, despite the sequence length remaining constant at ten-items per sequence, each RT profile shows sensitivity to the set compositionality, and varies according to the positions of the category boundaries. In the case of the first condition [AAAABBBCCC], see also fig. 4.5, although a number of increased RTs may be seen ($F_{9,198}$ = 37.7, $p<.0001$), only those occuring at the fifth, eighth and last positions were significant (all at $p<.05$, the former two

representing the only category boundaries of the array).

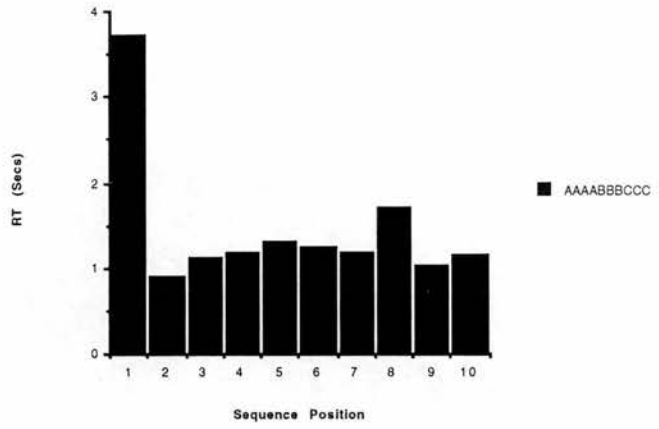


Figure 4.5 Mean RT profile for successful 10-item sequencing of three categories, each with three or four exemplars [AAAABBBCCC] (Ollie)

For the second condition [AAABBBBCCC], fig. 4.6, a number of increased RTs may again be seen ($F_{9,198}$ = 62.2, p<.0001), but this time only those occuring at the fourth, eighth and last positions were significant (all at p<.05, the first two representing the category boundaries of the array).

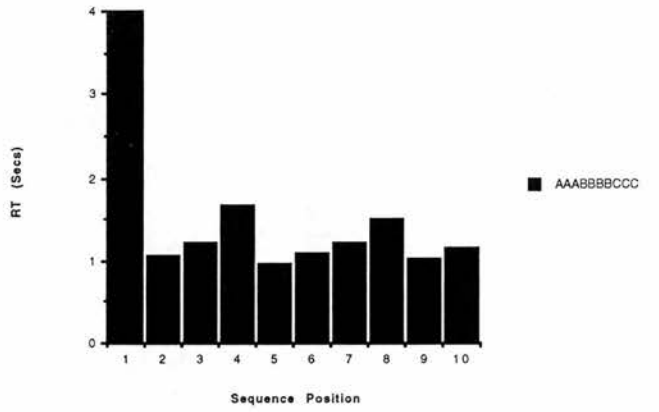


Figure 4.6 Mean RT profile for successful 10-item sequencing of three categories, each with three or four exemplars [AAABBBBCCC] (Ollie)

In the third and last condition of this phase [AAABBBCCCC], see fig. 4.7, the same subject may be seen to consistently produce increasing RTs at yet different positions ($F_{9,198}$ = 15.4, p<.0005) This time only those occuring at the fourth and seventh positions were significant (both at p<.05) where again each may be found to lay at the only category boundaries of the condition array.

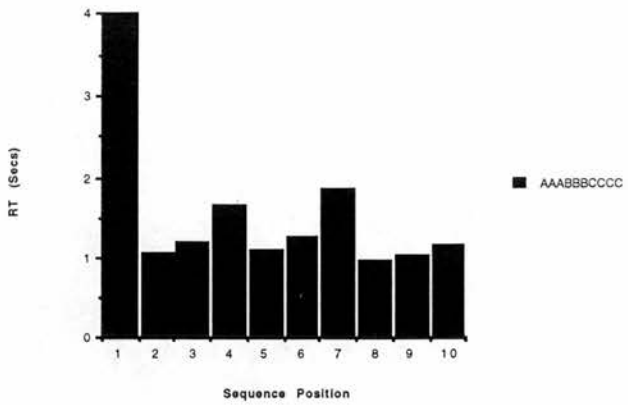


Figure 4.7 Mean RT profile for successful 10-item sequencing of three categories, each with three or four exemplars [AAABBBCCCC] (Ollie)

Discussion. Condition criteria were met by all of the monkeys in this phase, successfully reporting any one of three interspersed ten-item mixed triplet and quadruplet composed strings. Successful sequencing was upheld independent of compositionality using all three of the category markers with each maintaining their relative sequence order as presented in the previous phases. It can also be seen that there was once more an order of completion effect for each subject. All of the monkeys completed the [AAABBBCCCC] condition first, with the fewest number of trials and percentage errors to criterion, indicative of a generalisation effect - this particular condition differing from the previous nine-item training condition only in that an additional (equivalent) item needs be added in the last position. The other two conditions also resulted from single-item extensions of the same sequence as the first, but differed in that their extra icons were to be 'inserted' into ordinal positions necessarily internal to the previously learned sequence positions. It was quite clear from the examination of the same monkey's RT profile for the differing interspersed condition compositions, that maintenance of the pausing at category boundaries continued independent of the sequence position *per se*. The overall time taken to complete the ten-item sequences did not significantly differ between conditions, and the findings indicated that not only was compositionality a major factor in determining the nature of the RT profile for such sequences, but that the contour profile produced appears to be sensitive to the nature of the category boundary features belonging to the particular condition being interrogated on-line.

If the RT profiles were to reflect merely sequence position *per se* for the first and subsequent touches up until the end of a given string-length, one would not expect to record any differences relating to the composition of a sequence of identical length. However, post-hoc pairwise comparisons for adjacent touches revealed significant differences only at category boundaries (excepting the last position for the first two conditions) thereby proviving good evidence for the existence and use of some strategic planning (or at least economic management) with these large sequences. Not only may this be derived from an analysis of RT data, but such is further reflected in the persistent pattern and number of error-types shown by all of the subjects. The shift towards dominantly one-step forward looking errors was continuing, with more than 45% of one of the condition's error trials being accounted for by a single touch sequence despite the possibility of some twenty-four error types for each of the test conditions. Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner still remains unknown at this stage because, at present, any forwards-error touch immediately signalled the end of the current trial. However, it was now certain that the monkeys were quite capable of demonstrating a remarkably robust level of serial order control with three different ten-item sequences under supervised-learning conditions within which some degree of free search was permissible.

Phase 10

Rationale. Having by now satisfactorily completed nine- and ten-item sequences (whether composed of three or four exemplars of three different classes), the monkeys had now qualified for a further and final increase in sequence length. No indications were as yet fourthcoming as to what the upper limits of string-length for the demonstration of serial order control still might be for the monkey, and so a final phase comprising a twelve-item set was to be presented. At first offered cold for the purposes of generalisation/transfer evaluation, this phase would further inform us as to whether any of the monkeys would be able to continue differentiating between the components of such large stimulus arrays, and indeed be capable of executing a twelve-item sequence at all, let alone proceed to demonstrate such in a systematically controlled or principled way.

Subjects. All six adult monkeys took part in this phase.

Design and Procedure. Extending each monkey's core [ABC] sequence further again, this single condition probe phase introduced the monkey to a twelve-item set for the first time, cold. Adding a fourth exemplar to each of the existing categories [A], [B] and [C], the new sequence was composed of three categories, each with four identical exemplars [AAAABBBBCCCC]. Each subjects' task was to sequentially interrogate and touch all of its category [A] icons first, followed by all of the [B], then finally all of the [C]. There was no experimental criterion for success with this probe phase which was to be presented for two consecutive sessions according to the subjects disposition to the task, each presenting a maximum of fifty trials per session.

Results. All six monkeys completed two sessions (a total of 100 trials each) and all succeeded in reporting some correct sequencing for this twelve-item condition despite the much higher level of error possible with the newly extended string-length. The number of trials, errors and percentage errors (without phase criteria) for this condition were as shown in Table 4.37.

Table 4.37. Total numbers of probe trials, errors and % errors [AAAABBBBCCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	100	72	72.0%
Alfie	100	87	87.0%
Kissy	100	96	96.0%
Mimi	100	83	83.0%
Luba	100	96	96.0%
Ollie	100	59	59.0%

Although accounting for between 59% and 96% of the total trials for any one subject, it was found that despite there now being thirty possible error paths executable, not only were many of the error typologies under-represented, but more than 63% of the total number of error trials could be accounted for again by only two of them. For pooled data (N=6) the dominant error typology reported by the monkey was [AAAB] (43.0%), followed by [AAAABBBC] (20.8%), all of the nearest contenders attracted

less than 4.0% each. Due to the low numbers of correct trials with the twelve-item sequences at this stage, it was not possible to reliably use the RT data for useful analysis at this stage (but see results for Phase 11, a re-run of the current phase with a set criterion, below).

Discussion. Although there were no experimental condition criteria for this phase, all of the six monkeys succeeded in reporting some correct trials of a twelve-item string despite the much higher level of difficulty and performance required. This has provided the first demonstration of the abilities of a non-human primate to include the reporting of any sequence of this length, let alone a principled successful execution of a simultaneous three category, four-exemplar sequence, as was presented in the stimuli array for this condition. Despite the relatively low number of trials presented to each monkey in this phase, an indication of their successfully exploiting the principle of the task's simple classificatory scheme was to be seen in the error profile for incorrect trials. First touch errors, which with a nine-item array one might expect to see reported more frequently now, were recorded for less than 2% of the total error space for all subjects. Given the possibility of as many as thirty error typologies for this new twelve-item probe condition, error profiles revealed clear preferences for sequential touching in accordance with target category sequencing once more, with one-step forwards-errors [AAAB] and [AAAABBBC] accounting for more than 63% of all the error trials recorded (at a rate of 43% and 21% respectively). Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner still remains unknown at this stage due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (any forwards-error touch ended the trial). Evidence for the existence and use of any strategic planning (or at least economic management) with movement though this much larger sequence comes only from an analysis of error frequency. RT profiles produced for the very few correct trials did, however, show that much of the time taken to successfully complete the trial continued to be represented in the initial reaction time to the first touch (IRT), indicative of careful inspection and forward planning prior to sequence execution. Nonetheless, these findings do provide the first existence proof of the successful execution of a twelve-item sequence by a non-human primate (or any other species for that matter) and indicate a level of

performance consistent with the presence of classificatory processes in the monkey, at least in a supervised laboratory environment.

Phase 11

Rationale. Although reaching an experimental criterion for this stage would not be a prerequisite for continuation with the experimental series (that already being satisfied by the completion of the nine-item sequence as discussed above), Phase 11 was the last of the [ABC] acquisition phases of the study. Given the monkey's qualification for entry into the subsequent phases which explore a greater depth of search and possible hierarchical management of more complex strings, the aim here was merely to inform us as to whether any of the monkeys would be capable of consistently reporting a novel twelve-item sequence in a controlled and principled way. Such a competence remains well beyond the reported animal learning or seriation literature for a non-human subject of any species.

Subjects. Due to the presence of youngsters newly born to both Luba and Kissy (once more) in the colony rooms, this phase was presented only to three of the monkeys (Charlie, Alfie and Mimi). Mimi was also occasionally carrying her last born infant (by now independently mobile) with her whilst in the testing situation.

Design and Procedure. Continuing with the extending of the sequence length of a string of familiar icons for each monkey, this last of the [ABC] acquisition phases presented arrays of a novel twelve-item sequence, [AAAABBBBCCCC]. Interspersed trials of control nine-item [AAABBBCCC] string were also shown periodically to assist the interpretation of any failure to demonstrate or improve performance with the twelve-item test condition. As for the previous phases, each subjects' task was to sequentially interrogate and touch all of its category [A] icons first, followed by all of the [B], then finally all of the third category [C]. The two conditions were presented at a 1:4 ratio for triplet and quadruplet sequences respectively, employing multiple exemplars of the same icons of categories [A], [B] and [C] as previously used. Following every twenty quadruplet string exposures, maintenance levels of three out of four successful control trials were also required in order that the test condition continue to be presented. The phase was completed when the quadruplet

condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three of the monkeys completed this final phase test condition by achieving the 75% criterion level required, whilst simultaneously maintaining a high standard of performance with the control condition. For the twelve-item sequence condition, the mean number of trials, errors and percentage errors for each of the subjects were as shown in Table 4.38.

Table 4.38 Nos. of trials, errors and % errors to criterion [AAAABBBBCCCCC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	569	360	63.3%
Alfie	3081	2234	72.2%
Mimi	1104	817	74.0%

Analysis of the incorrect trials revealed a continuation of the pattern found previously, in which unequal representation of the possible error types was evident. Accounting for between 63% and 74% of the total test trials completed for any one subject in this phase, it was found that despite there now being thirty possible error paths executable, although all of the error typologies were represented, more than 59% of the total number of error trials could be accounted for by only two of them. For pooled data (N=3) the dominant error typology reported by all monkeys was [AAAB] (40.9%), followed by [AAAABBBC] (18.4%), the nearest contenders attracted less than 9.0% [AAAABBBBCCCC]. This latter figure is relatively high in comparison to previous sequence error profiles, and for the first time suggests that a degree of strain is being imposed on the subject. This length of the sequence to be executed (12 items) with four exemplars of each class produced significant number of reiteration errors occurring in the final position, mostly realised as a reiterative touch to a [C] class icon (a total of 301 trials) compared to reiterative touches to either an [A] or a [C] (totalling 3 and 17 trials respectively).

Although confined to the analysis of only two monkey's data (the mother Mimi's RT profile being confounded by her infant's attendance in the test environment), reaction time data for correct trials reveals a different

overall pattern of responding from the differences seen previously in the nine-item phase. Although the pooled results well resemble the previous profiles found, Alfie's individual RT profile remained primarily linked to a categorical boundary effect (with significant ($p<.05$) pausing at the beginning of each category boundary) but also showing significant rises at the third and ninth positions (see fig. 4.8). Charlie who showed a markedly *reduced* time to be taken touch the first icon of the last category, showed significantly increased RTs for the fourth and fifth icons, the latter between the first and second category boundary (see fig. 4.8). Further, whilst all analysis of variance were significant beyond the 1% level, this was the fisrt time that both within and between category RT values showed other significant intercategory exemplar effects.

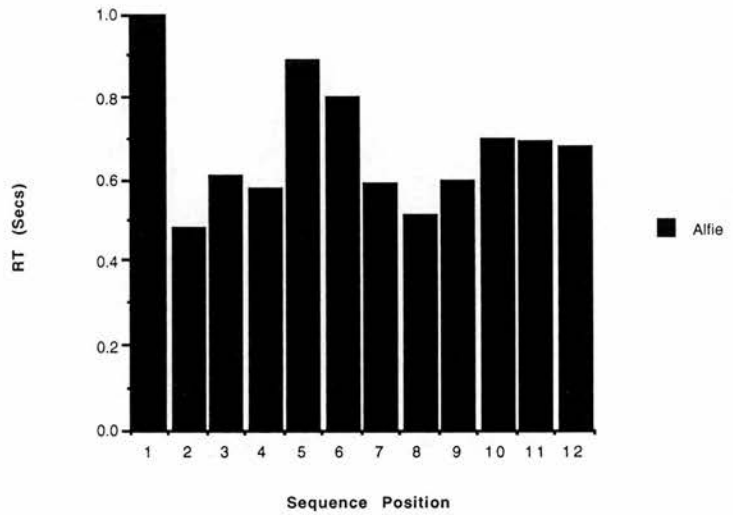


Figure 4.8 Mean RT profile for successful 12-item sequencing of three categories, each with four exemplars, in a Cebus monkey (Alfie)

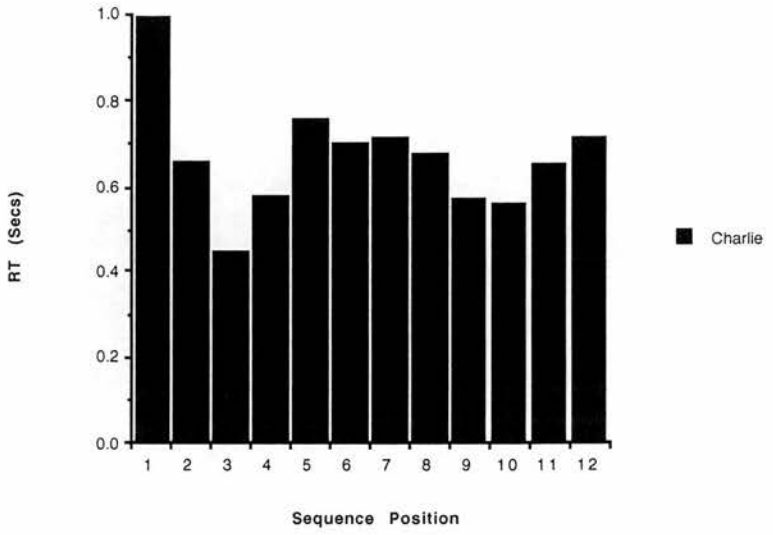


Figure 4.9 Mean RT profile for successful 12-item sequencing of three categories, each with four exemplars, in a Cebus monkey (Charlie)

Discussion (and interim summary: phases 9-11). Condition criteria were met by each of the three monkeys in this last incremental set phase, their successfully reporting both nine- and twelve item triplet and quadruplet sequences. This success was consistently demonstrated despite the enormous demands of the task, using three category markers with each maintaining their relative sequence order as presented in their previous [ABC] training phase although with different absolute serial positions. Another measure of their achievement here, is that there was no decrease in their degree of perseveration despite the much increased cost of making an error the further down the sequence the monkey had progressed for any one trial. First touch errors were recorded now less than 1% of the time (except as recorded for Mimi, who's particularly high rate (15%) is explained by the presence of an occasional third hand on the screen !) indicating a high degree of inspection of the now greatly expanded array despite the greatly decreased initial reaction times. Furtherstill, with the possibility of now of thirty error typologies for the test condition (and a staggering 1 in 10^9 chance of getting it correct), error profiles revealed clear preferences for touching in accordance with the previous [AB] and [BC] category sequencing once more, with two forward-error types accounting, on average, more than 60% of all the error trials for the three-quadruplet condition. The monkey could have

proceeded to produce a successful exhaustive search of the array having effected repair 'down the line' but this at present remains unknown due to the nature of the machine constraints imposed upon the subject (a forwards-error touch ending the trial).

It has now become quite clear that all of the monkeys so far tested (see table 4.39) were quite capable of demonstrating consistent seriation skill with sequences of up to, and including, a twelve-item string length under supervised-learning conditions for which a degree of free search is permissible. Whether this species could tolerate any further extension of their existing series' [ABC] icon sets in terms of breadth (say, to fifteen or twenty items) is currently not known, but there is little to suggest that any subject is anywhere close to reaching the upper limits of performance at either their nine- or twelve-item current stages.

Table 4.39 Numbers of monkey subjects completing sequences to 12-item string-lengths to date (100%).

3-items	6-items	9-items	12-items
6	6	6	3

Indeed, at least one monkey (Charlie) had completed his twelve-item set in a much shorter time (and with fewer trials) than was taken to satisfy the criteria levels of performance for his own nine-item sequence, suggesting a reduced level of effort being required to solve the larger set sizes. Evidence for the existence and use of some strategic planning (or at least economic management) with this larger sequence comes not only from an analysis of error frequency, but is supported by the finding of differentiable RT profiles especially for the six-, and ten-item sets, the nine- and twelve-item sets indicative of salient chunking, although not as yet exclusively repartitioned for differential compositionalities. If the RT profiles were to reflect merely sequence item number selection from the screen array, one would not expect to see any differences necessarily relating to the composition of the string. However, the findings indicate not only that compositionality is indeed a factor in determining the nature of the RT profile for sequence positions, but that the actual profile contour produced, appears to reflects the nature of the categorical boundary features of the condition being interrogated (see also summary table 4.40).

Table 4.40 Monkey subjects showing significant latency rise-times only at the category boundaries

Subject	AABBCC		AAABBBCCC		AAAABBBBCCCC	
	1st	2nd	1st	2nd	1st	2nd
Charlie	x	◊	◊	◊	◊	x *+3,11,12*
Alfie	◊	x	◊	◊	◊	◊ *+3,10*
Kissy	◊	x	◊	◊	-	
Mimi	x	◊	◊	◊	-	
Luba	◊	◊	◊	◊ *+8,9*	-	
Ollie	x	◊	◊	◊ *+5*	-	

◊= ($p<0.05$); x = NS; *+n,..*= other sig. ($p<0.05$) RT position (s).

These findings provide further convincing evidence for spontaneous control of classificatory processes extant in the monkey. Such phrasing effects, in which the search of items within a class is generally much faster than for that of the first item of a class, is akin to a form of chunking, and offers a further indication of the ongoing development of classificatory expertise in the monkey. Continuing with our voyage of discovery, we now knew that the monkeys would engage with manageable set sizes of sufficient length so as to allow their further differentiation of the same icons to provide a second level of classification (see fig. 3.3). The next series of investigations were designed explicitly to explore the monkeys' tolerance of feature variation and the expansion of their core-[ABC] icons in terms of an increase in the depth of search, rather than continuing to expand in breadth.

(B) [ABC] Equivalence Phases

Introduction

Having now achieved consistently successful performances with all monkey subjects at at least the nine-item, three category classification level, rather than continuing to increase the number of exemplars in each category (resulting in a search problem of its own) it was decided at this stage to explore the *depth* of search that the monkey might be expected to report. One way of decreasing the memory load being placed upon a subject working with very long sequences, would be for them to attend to the presence of any distinguishing features salient to the individual items to be seriated, which might then be used in some hierarchically organised way. The problem could thereby be solved by

selectively reducing the larger search space into more manageable chunks (as one might often do when trying to remember and dial telephone numbers). Following the completion of their last nine- or twelve-item string phase, all monkeys were therefore to return to their nine-item string lengths [AAABBBCCC] and be then exposed to two forms of physical variation, at this stage to be presented as equivalence phases. Although it might have been eventually possible that the monkeys might have at this time proceeded to spontaneously report principled ordering of icon sets with categorical exemplars varying in colour and/or size under supervised conditions without much difficulty, at this stage of our 'voyage of discovery' it was very important to first establish the degree of the monkeys' tolerance to such variation in the appearance of their existing icon sets. In order therefore to ascertain whether the feature variations to be used for each subject in the future were to be acceptable to the monkey, a probe set of equivalence phases was undertaken by each. At this stage there would be no requirement for the subject to explicitly order the new nine-item sets according to any particular size or colour code (although the subject could self-impose such a constraint), the aim was simply to see whether the monkey would maintain its existing levels of performance with a nine-item, three category set whilst at the same time tolerate feature variation within each category. Such steps are important, because without the knowledge of the monkey's ability to distinguish them, let alone be able to view them as being class-equivalent, one would not be in a position to readily interpret their performance with an explicitly hierarchical search task, especially in the face of their failure (producing a large amount of initial resistance, or increases in the numbers or types of errors being demonstrated). The [ABC] equivalence phases 12 through 13 to be employed are shown in Table 4.41 below and include the details of interspersed conditions and completion criteria for each phase. All subjects participating in this experimental series started with Phase 12. Each phase required an extremely high level of vigilance and consistently high levels of performance to be maintained (with thirty possible error typologies being possessed by each). Notwithstanding icon differentiation problems these are all, in principle, equally difficult tasks.

It is however notable at this point, that a difference could possible be expected to emerge as the result of the natures of the two equivalence dimensions being presented, each task providing different opportunities

Table 4.41 Equivalence Phases 12-14 [ABC]

Phase No.	Condition	Completion criterion
12	AAABBBCCC - monochrome	15/20
13	AAABBBCCC - monochrome	3/4
(colour)	$A_CA_CA_CB_CB_CB_CC_CC_CC_C$	15/20
14	AAABBBCCC - monochrome	3/4
(size)	$A_SA_SA_SB_SB_SB_SC_SC_SC_S$	15/20

for observing their categorical exemplar's connectivities. Whereas for the colour equivalent categories a necessarily arbitrary basis will always be employed in the executed sequences (although they might become held consistent across categories), the size-equivalence stimuli sets offer clear relational intra-class variation in that they may be touched in order of increasing of decreasing size for any given category. For each of the equivalence phases, individual monkeys were to be self-selecting by their continuing successful attainment of the given phase criteria as they progressed through the tasks. Phases 13 and 14 were given as assigned first or second on a randomly distributed basis following completion of Phase 12, so as to control for any order effects of equivalence-task experience. There was no explicit tutoring of the monkey for single unique solutions in any of the equivalence conditions, and for each of the phases 12 through 14, only [A..B..C..] category order was required to be maintained.

Stimuli

The actual sequence compositions presented during these phases were drawn from a larger number of possible decomposition strings derivable from an expandable [ABC] icon set, but are quite conservative in their realisation in accordance with the examples given in fig. 3.3. As a general code, the following key provides examples of the nomenclature to be used in identifying string composition throughout the following tables:

A= 1st category monochrome contour shape (e.g., white outline star)

B= 2nd category monochrome contour shape (e.g., white outline hexagon)

C= 3rd category monochrome contour shape (e.g., white outline square)

A_C= 1st category shape free to vary in colour (e.g., green, white or white contour star)

B_S= 2nd category shape free to vary in size (small, medium or large hexagon)

For each of the colour equivalence condition stimuli, the size remained invariant and of the same dimensions as all those seen in the previous phases. For the size variation condition, the medium size was of the same familiar shape and dimensions, but with the smaller and larger stimuli showing geometrically scalar changes in their surface areas according to the halving or doubling of the central linear axis of the medium-sized icon shape.

Phase 12

Rationale. In order that the ensuing equivalence phases be consistently interpretable for both colour and size variations, it was important that the baseline nine-item set be devoid of colour cues from the outset. In order to achieve this, the existing [ABC] categories' shapes as previously used in the nine-item sequence [AAABBBCCC] were maintained, but following its colour feature subtraction, was then presented as monochrome contoured (familiar) shapes for the base-line comparison (e.g., white star outline on a black background). The [AAABBBCCC]-monochrome contour stimuli set was also to form the base-line condition for the equivalence phases which were to follow. The subject's performance on this phase would also inform us as to whether the previous characteristics of successful seriation might be due merely to familiarity effects related to the particular stimuli employed, or indeed to the supervisory nature of the incremental technique used to produce these long sequences in the first place. Bearing in mind that although the subjects were now two to three years into the experimental series, his would be the first time that any monkey was to view a novel nine-item array containing previously unseen icons for a serial, exhaustive search task.

Subjects. Five adult monkeys took part in this phase. One of the monkeys (Mimi) was heavily pregnant, and another female (Ollie) was carrying her newly born infant. Both were allowed to work for as long as they continued to present for transfer to the experimental test-room.

Design and Procedure. This new nine-item [AAABBBCCC]-monochrome contour base-line condition was given without a confounding control, providing a singly consolidated criterion for three familiar categories (at

least by shape). As for all of the previous phases, each subjects' task was to sequentially interrogate and touch all of its category [A] icons first, followed by all of the [B], then finally all of the third category [C]. For any given trial, the three icons of each category were only distinguishable by their spatial location on the touch screen. An example of screen array for this phase is shown in figure 4B.4. This single condition presented error-types covering twenty-one possible incorrect paths with its nine-item string length comprising three categories, each with three physically equivalent monochromatic icons The phase was continued until performance levels met the usual task criterion of fifteen correct out of the last (cumulative) twenty trials.

Results. All five monkeys achieved the experimental criterion level set for the new physical equivalence condition, and did so in far fewer trials than that used to have successfully completed their original nine-item sequences. The mean number of trials, errors and percentage errors to criteria for the [AAABBBCCC-monochrome contour] conditions is shown in Table 4.42 below.

Table 4.42 Nos. of trials, errors and % errors to criterion [AAABBBCCC Monochrome Contours]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	766	478	62.4%
Alfie	878	555	63.2%
Luba	164	80	48.8%
Mimi	307	230	74.9%
Ollie	91	37	40.6%

Although there were twenty-one possible error-types for this condition, for all subjects, more than 85% of all errors could be accounted for by two forwards-error types. Indeed, for two of the subjects more than 75% could be accounted for by only one. The possible errors which could have occurred in sequence positions 1, 4, 7 and 8 (fifteen in total) were hardly represented at all in the monkey's error profile, and zero values were often recorded. Both of the dominant error-types had in common that they represented one-step, feed-forward errors at the category boundaries [AAB] and [AAABBC] in the third and sixth positions.

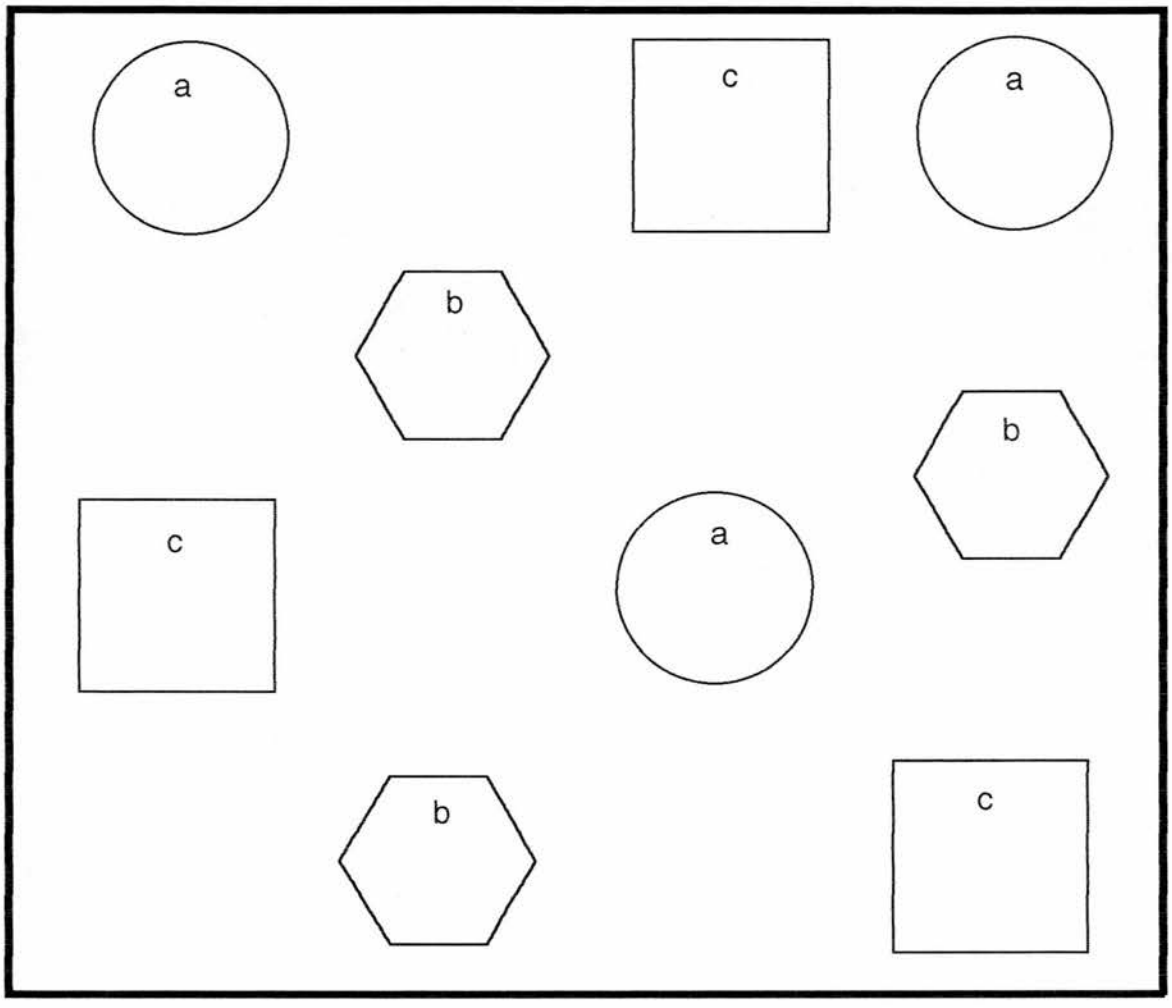


Figure B.4 Example of screen array for nine-item set: [AAABBBCCC]- monochrome contours (Phase 12)

Category boundary effects were also in evidence following analysis of pooled reaction time profiles (see figure 4.10) which showed increasing times to be taken at both the [AB] and [BC] boundaries only ($p<0.002$). For individual animal profiles, all category boundary effects were significant and positive ($p<0.01$) excepting one of Charlie, for which he showed a significantly decreasing RT ($p<0.002$) for his second category [BC] boundary (although his [AB] transition was the same as the others as might otherwise be suggested by the pooled data illustrated in figure 4.10 below. Other post-hoc evaluations revealed either decreasing RTs or no significant differences at all for adjacent pairwise comparisons.

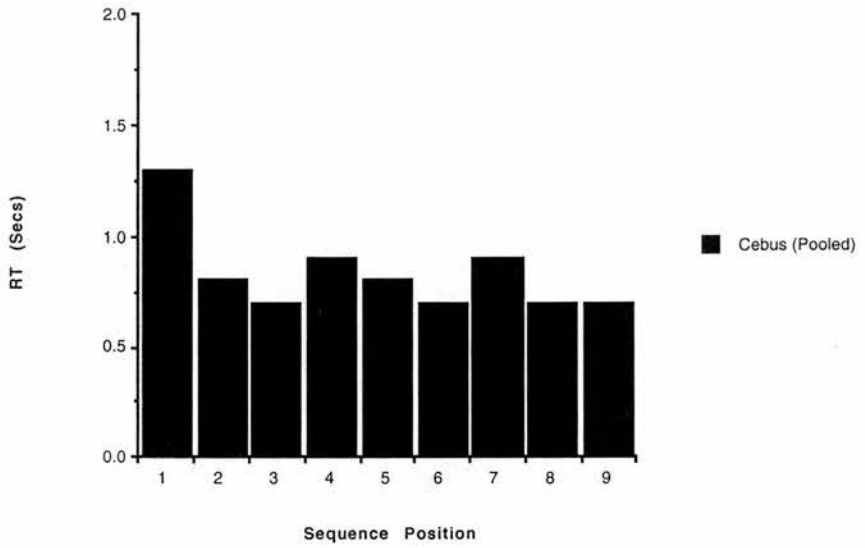


Figure 4.10 Mean RT profile for successful 9-item sequencing of three categories, each with three identical exemplars, in Cebus monkeys (N=5, Pooled)

Discussion. This phase has revealed for the very first time, the ability of the monkey to successfully seriate a nine-item set composed of novel icons, the absolute properties of which they had never before seen. Although preservative of outline shape, the stimuli of the array were similar to those of each monkey's previous sets, only having undergone cue subtractivity, in that each of the three category icons now had both their colours and shading density removed. Despite this loss of stimuli properties, all subjects showed fast uptake and sensitivity to the remaining icon features, and in comparison to the original nine-item [AAABBBCCC] acquisition data, the new arrays were successfully manipulated with up to 75% fewer error trials being demonstrated before

reaching the phase criterion for completion. Likewise, the total number of trials to criterion were also markedly less (in the case of one monkey, from previously more than a thousand trials to this time less than a hundred) with all subjects requiring fewer than 50% of the number of trials in order to have met their criterion-run than did they for the earlier coloured-icon seriations. The actual percentage of incorrect trials to criterion in this phase also fell for every animal. Error types data was also very informative with more than half the counts remaining at zero, and almost all of the errors committed becoming exclusively accounted for by a one-step forwards-error seen at either one of the category boundaries. Of these, more were occurring at the later [BC] boundary, and, excepting the case of Charlie, this was also indicated to be a significant ordinal position for successful trials. As indicated by the pooled RT profile, the inspection time for the first touch of each new category icon in the sequence was significantly longer than were the times for the intra-category touches. Furthermore, the absolute time from the beginning of the trial to first touch was much reduced, suggesting that the subject is now able to 'chunk' the larger array into smaller bites relatively quickly, perhaps making better predictions concerning the intra-class touches (based upon location identification only) once the salient common icon feature had been identified. Clearly indicative of good acquisition skill transfer and clear generalisation of class inclusion ability, this first test of a monkey's strategic competence with cold serial order control of such a large list has been very enlightening. The monkeys were now starting to inform us that they were themselves coming to be quite well task-informed.

Phase 13

Rationale. Having now demonstrated their first successes with a novel nine-item sequence with physical equivalence characteristics, it was now possible to explore the degree to which the monkey might tolerate feature variation within a class whilst maintaining categorical seriation. In order that the animal might inform us of their abilities to spontaneously classify in a more free-search situation, it was first necessary for us to determine whether they might report class inclusion for 'known' category icons which might vary along some experimentally manipulated dimension (for example, colour or size) under supervised learning conditions.

Subjects. Five adult monkeys took part in this phase. One of the females (Mimi) was pregnant and very close to term, but she was allowed to work for as long as she continued to present herself for transfer to the experimental test-room. After giving birth to her offspring in the colony room, Mimi continued to present and join the daily experimental sessions (complete with ventral clinging infant) but often did not complete many trials. It is also noteworthy here, that her working lab-partner (Charlie) seemed to be also quite distracted (although not upset) by their presence and showed atypical attention and touch-screen behaviour for a while.

Design and Procedure. This first non-physical-equivalence phase presented another completely novel looking three category, three exemplar, nine-item sequence to be exhaustively interrogated as before, but now, and for the first time, comprised of icons that could each be uniquely identified within the array. Free to vary within each category by colour [$A_cA_cA_cB_cB_cB_cC_cC_cC_c$], the subject was nonetheless only required to report the sequence by category as before (i.e., touch all [A_c] before all [B_c] before all [C_c]). Although it was necessary for the monkey here to observe the equivalence of particular icons for the purposes of class inclusion, it was not necessary to touch particular icons in a fixed ordinal position within a category (although they could self-impose such a constraint upon themselves if they so wished). An example of screen array for this first colour equivalence condition is shown in figure 4B.5. Interspersed with these trials the subject was also required to maintain control levels of three out of four successful trials of a [AAABBBCCC-monochrome contour] control condition in order that the colour-equivalence test condition continue to be presented. The phase was completed when the equivalence condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

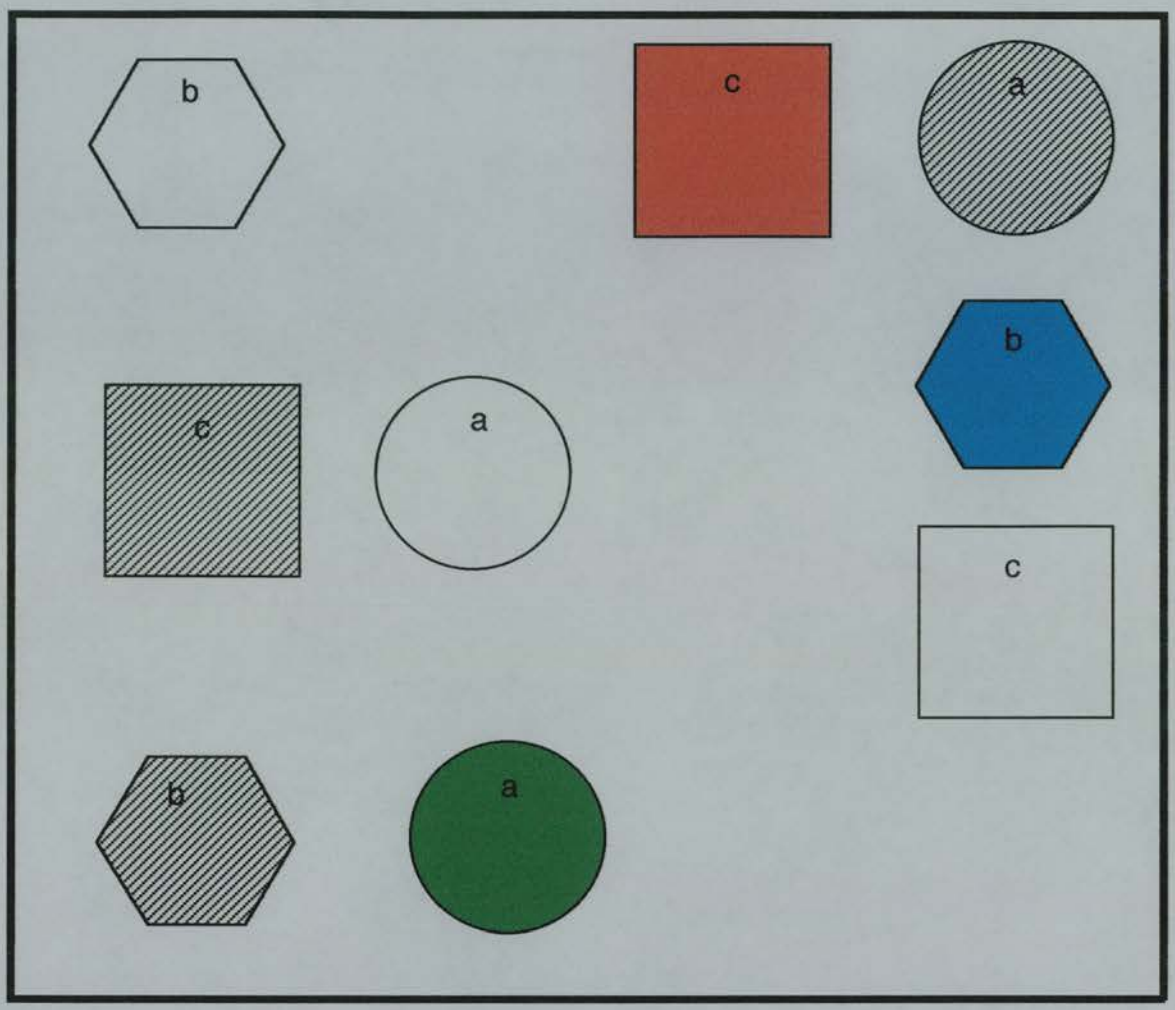


Figure B.5 Example of screen array for nine-item set:
[AAABBBCCC]-Colour Equivalence (Phase 13)

Results. All of the Cebus completed this 'colour equivalence' task whilst the 75% criterion required for the control condition were consistently upheld by all subjects for the duration of the phase. The mean number of trials, errors and percentage errors to criteria for the test condition were as shown in Table 4.43.

Table 4.43. Nos. of trials, errors and % errors to criterion [AAABBBCCC Colour Equivalence]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	438	271	61.8%
Alfie	636	468	73.6%
Luba	354	243	68.6%
Mimi	1346	938	69.7%
Ollie	465	262	56.3%

Although there were twenty-one possible error-types for this condition, for all subjects, more than 78% of all errors could be accounted for by only two forwards-error types. The pattern of errors was similar to that found in the the last phase, with both of the dominant error-types representing one-step, feed-forward errors at the category boundaries [$A_cA_cB_c$], and [$A_cA_cA_cB_cB_cC_c$], in the third and sixth positions respectively. Category boundary effects were also in evidence following analysis of pooled reaction time profiles (see figure 4.11) which showed significantly increased times to touch only at the [A_cB_c] and [B_cC_c] boundaries ($p<0.01$).

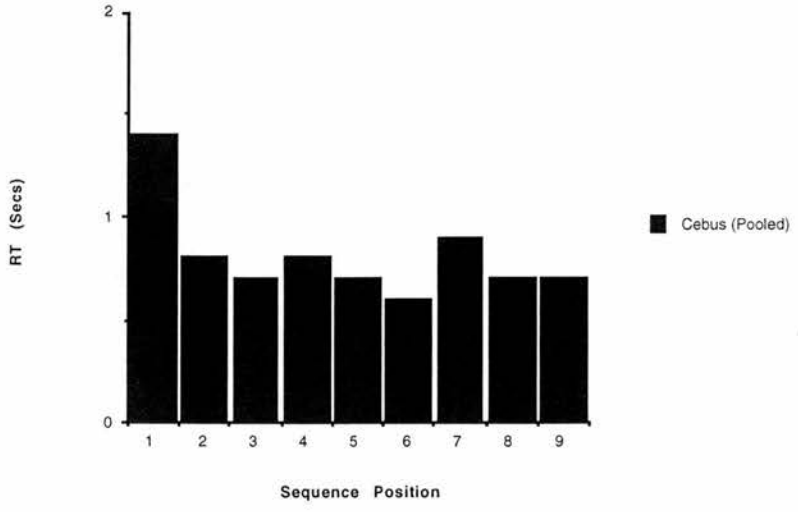


Figure 4.11 Mean RT profile for successful 9-item sequencing of three categories, each with three unique exemplars (colour variance), in Cebus monkeys (N=5, Pooled)

For individual animal profiles, although all analyses of variance were significant ($p<0.01$), pairwise reaction time data at category boundaries effects were not significant in all cases. For Charlie, no significant difference was found at the [B_cC_c] position (which still showed a decreasing RT value), and for Alfie, neither boundary figures yielded results of significance below the 5% level (although directionally consistent with the other three monkeys). Following analyses of the actual touch sequences for correct trials, there was no significant effect of spontaneous ordering of icons within a category. The only reasonably consistent findings was for 'end-anchor' effects for which in the case of Ollie, there was a preference for starting her sequence with the original first icon of the previous icon set (preserved shape and colour, [A_c = A]), and for Mimi, who in contrast tended to end her sequence with a green cross.

Discussion. This phase has provided the first demonstration of non-human primates successfully seriating a nine-item set composed of three categories, within each of which there were three exemplars free to vary in their coloration characteristics. Although preservative of outline shape, the stimuli of the array were similar to those of each monkey's previous phase, but this time having undergone cue additivity, in that each of the three category's icons now had both colour and shading density variations. All subjects showed even faster uptakes and sensitivity to the icon features of their stimulus arrays, and in contrast to the previous novel nine-item [AAABBBCCC-monochrome contour] acquisition data, the new arrays were successfully manipulated with various patterns of trials and errors to criterion. Charlie and Alfie were very much quicker in reaching the phase' condition criteria than were they for the previous phase. The two nursing females took significantly longer, although still requiring less trials than they needed to complete their original [AAABBBCCC] nine-item set. Error types were very similar to those seen with the physical-equivalent string in the previous phase, with some extra representations being contributed by the attempts of the carried infants to interact with the touch-screen apparatus. Good acquisition skill transfer and clear manipulation of early class inclusion abilities were evident from the outset of the phase, in which all but one monkey (Mimi) saw at least one correct test trial in their very first session.

Phase 14

Rationale. Having seen the first successes with a novel nine-item sequence with colour equivalence characteristics, it was now possible to explore whether the monkey might tolerate any further feature variation within the same classes for the same categorical seriation task. Again, in order that the animal might inform us of their abilities to spontaneously classify in a more free-search situation, it was first necessary for us to determine whether they might continue to report class inclusion for their 'known' category icons free to vary, this time, in size, under supervised learning conditions.

Subjects. Five adult monkeys took part in this phase. One of the females (Mimi) had recently given birth and continued to present and join the daily experimental sessions (bringing her infant along to work with her) but often did not complete her sessions. Another female (Luba) was also now in the later stages of an unexpected pregnancy.

Design and Procedure. This second of the non-physical-equivalence phases presented another novel three category, three exemplar, nine-item sequence to be exhaustively interrogated as before, but on this occasion, and for the first time, comprised icons that could again each be uniquely identified within the array. Free to vary within each category, this time by size [$A_S A_S A_S B_S B_S B_S C_S C_S C_S$] the subject was required to report the sequence as before (i.e., touch all [A_C] before all [B_C] before all [C_C]). So although there was need for the monkey to observe the equivalence of particular icons for the purposes of class inclusion, for this phase, it was not necessary to touch particular icons in a fixed ordinal position within a category (although they could self-impose such a constraint upon themselves if they so wished). An example of screen array for this first three-size equivalence condition is shown in figure 4B.6. Interspersed with these trials the subject was also required to maintain control levels of three out of four successful trials of a [AAABBBCCC-monochrome contour] control condition in order that the size-equivalence test condition continue to be presented. The phase was completed when the equivalence condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

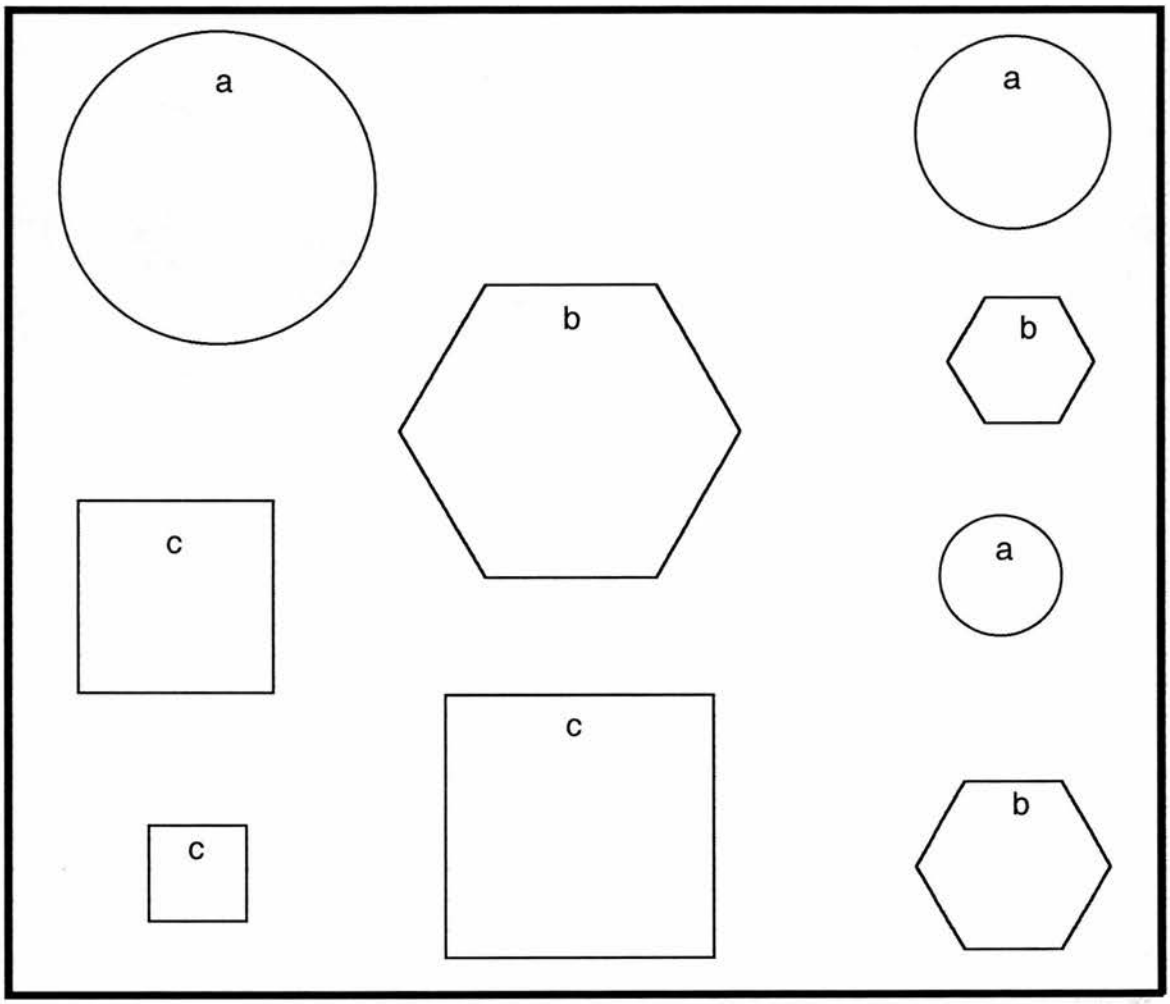


Figure B.6 Example of screen array for nine-item set: [AAABBBCCC]-Size equivalence (Phase 14)

Results. Only four of the five monkeys achieved the experimental criterion level for the new size-equivalence condition, and did so, again, in far fewer trials than were used to successfully complete their original nine-item sequences. All four subjects continued to the uphold a 75% criterion maintenance performance for the control condition throughout the duration of the phase. The mean number of trials, errors, and percentage errors to criteria for the size-equivalence condition [$A_S A_S A_S B_S B_S B_S C_S C_S C_S$] is shown in Table 4.44 below.

Table 4.44 Nos. of trials, errors and % errors to criterion [AAABBBCCCC Size Equivalence]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	470	302	64.2%
Alfie	1361	808	59.3%
Luba	2167	1713	79.0%
Mimi	476	256	53.8%
Ollie	-	-	-

Although there were twenty-one possible error-types for this condition, for all subjects, more than 82% of all errors were accountable for by only two forwards-error types. Both of the dominant error-types were single-step, feed-forwards errors at the category boundaries [$A_S A_S B_S$] and [$A_S A_S A_S B_S B_S C_S$] in the third and sixth positions. Category boundary effects were also in evidence (in different directions), but data could not be pooled this time due to the unequal variance shown between the monkey's RT data sets. Analysis of individual monkey reaction time profiles show Charlie (see figure 4.12) with increasing touch times, but not only the [$A_S B_S$] and [$B_S C_S$] boundaries. Indeed, these effects were found not to include any statistically significant increases in reaction time, although the subsequent decreases were ($p<0.01$).

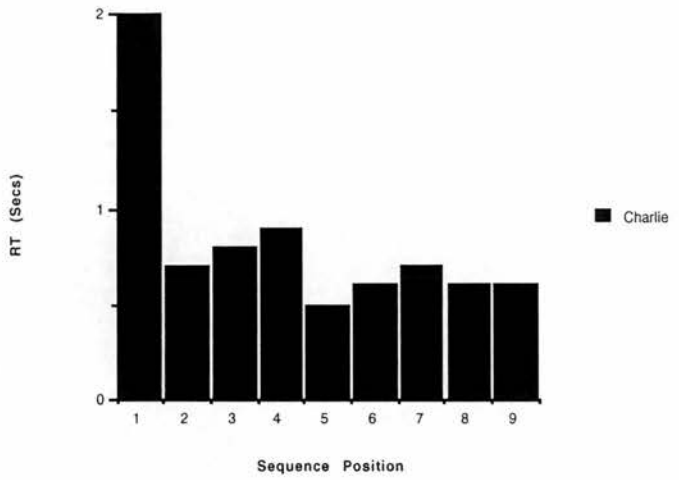


Figure 4.12 Mean RT profile for successful 9-item sequencing of three categories, each with three unique exemplars (size variance), in a Cebus monkey (Charlie)

A quite different profile was to emerge from Alfie, who showed significant effects at both the $[A_S B_S]$ and $[B_S C_S]$ boundaries ($p<0.05$). However, the novelty here was that each was in a different direction, as may be seen in figure 4.13. Other adjacent pairwise evaluations which showed significant effects, occurred at positions six and nine, although consistent touch data correlated with these findings other than his preference for touching the largest of each category icon first on these occasions.

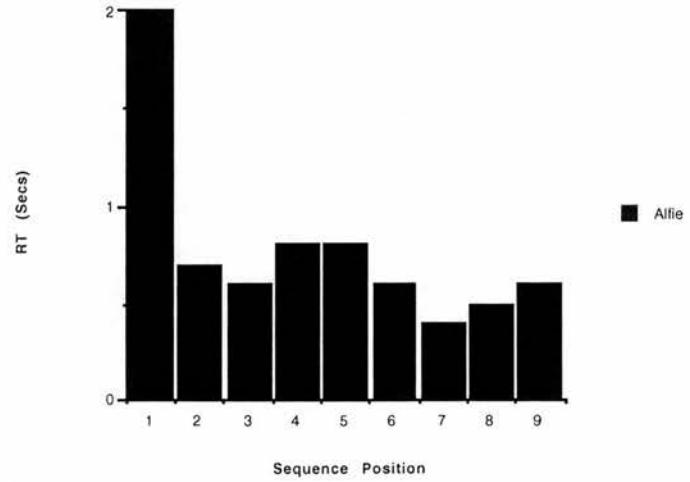


Figure 4.13 Mean RT profile for successful 9-item sequencing of three categories, each with three unique exemplars (size variance), in a Cebus monkey (Alfie)

In the case of Mimi (see fig. 4.14), we see for the very first time a significant decrease in her time to touch the first item of the second

category ($p<0.005$), itself also an unusually fast response linked to icon preference (always touching the largest first, rather than merely reflecting a much longer search for the last item of the previous category). Note was also made of a much reduced (50%) start up time.

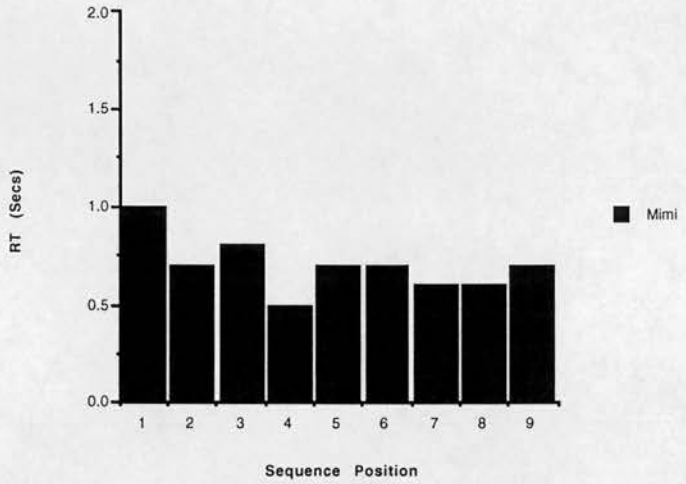


Figure 4.14 Mean RT profile for successful 9-item sequencing of three categories, each with three unique exemplars (size variance), in a Cebus monkey (Mimi)

Luba's reaction time data was not included for the purposes of analysis due to its lack of consistently representative sampling availability. Analyses of the actual touch sequences for correct trials in the final criterion-run, revealed some effect of idiosyncratic icon ordering tendencies. Although specific transitional probabilities were not calculated, it is noteworthy here that for Charlie, five out of his last ten touch sequences began with a fixed ordering of the first three items of the first [$A_S A_S A_S$] category. Furthermore, this sequence was a self-imposed production of a serial monotonic ordering by increasing size. His second category was likely to be started with the smallest exemplar (actually on 60% of trials for the same data set) but did not reliably continue in the same way. For the third category, the only significant observation was that Charlie would reliably choose to end his sequence with its smallest exemplar. In the case of Mimi, there were also indications of a tendency to relax into an ascending monotonic size series for her first category (40%) and would otherwise (40%) touch the largest first followed by the smallest, then the middle-sized. She showed no preference for ordering the second category at all, but did produce an unique profile for her third. In 8 out of her last 10 correct trials, Mimi was to place the middle-sized icon in the eighth

ordinal position. The only reasonably consistent pattern of sequence touching from Alfie's performance on size equivalence, was that he did seem to show a preference for starting each category of the sequence with its largest icon.

Discussion. This size-equivalence phase has provided the first ever demonstrations of non-human primates successfully seriating a nine-item set composed of three categories, within each of which are three exemplars free to vary in their size characteristics. Although preservative of outline shape, the stimuli of the array were similar to those of each monkey's previous phase, but underwent cue additivity, in that each of the three category's icons were individually discriminable from each other by size. Subjects showed variation in their degrees of initial uptake and sensitivity to the icon features of their stimulus arrays, and in contrast to the previous novel nine-item [$A_c A_c A_c B_c B_c B_c C_c C_c C_c$] acquisition data, the new arrays, although successfully manipulated by all monkeys, were completed faster by some and slower by others in terms of trials and errors criterion. Mimi was very much quicker in reaching the phase' condition criteria than she was for the previous (colour equivalence) phase. The abnormally high number of trials (and percentage error trials) to criteria for Luba may be explained by her post-natal circumstances. Luba was to show much task-attentional variation during her contribution to this phase (despite her persistently presenting herself for work !). Error types were very similar to those seen with the other equivalence conditions in Phases 12 and 13. Good acquisition skill transfer and clear manipulation of early class inclusion abilities were evident from the outset of the phase, in which all but one monkey (Charlie, this time) saw at least 25% correct test trials among his first twenty of session one. But perhaps the most noticeable finding of this phase was that the monkeys were seen to be self-imposing an order of completion for items within a category, so providing a solution to the seriation problem that was indicative of spontaneous serial order control within a task not requiring such explicit ordering. Further, in the case of two monkey subjects (Charlie and Mimi) there was very clear evidence for the demonstration of spontaneous, monotonic size seriation. In the case of Charlie, a preference for starting his sequence with a set of three icons of increasing size, and for Mimi, a more variable ordering pattern, in which a fixed middle-sized icon was placed in position eight, allowed both

decreasing and increasing size seriations to be demonstrated at the end of her sequences. This finding, in contrast to those of the colour-equivalence condition, was also to be reflected in the somewhat unusual reaction time profiles for each individual monkey. Most noticeable was the lack of increase in the times taken to touch the first icon of the second and third categories. One might suggest that at least for the case of Charlie, if the animal is already committed to identifying and touching the first category's icons in a particular order, the first item of the next category could become part of that initial search now that a reduction in operational memory has been effected. Likewise, in the case of Mimi, the loss of increased time, indeed decreasing time, taken to touch the first icon of the third category is also explainable by virtue of the need to only search for either of two icons (the smaller or larger of the $[C_S C_S C_S]$) when fixing the middle-sized item to the central ordinal position (although eighth ordinal position of the entire sequence) of the final category triad. This three-size condition completes the equivalence phases of the ongoing task series using the already sophistocated subject proficient with at least a nine-item series. Each monkey subject has now clearly demonstrated 9-item seriation, tolerating stimuli feature variation, in both size and colour.

Interim Summary (Phases 12-14)

What has been revealed after some two and one half years of the program, is that the *Cebus apella* monkey has proven capable of reporting categorical seriation of sequences comprising three categories, each with at least four exemplars. There is at present, however, no indication that this may be anywhere close to the upper limit of the possible string lengths negotiable by the monkey under supervised conditions. Further to this finding, evidence has been acquired to support the idea that, using non-linguistic tasks of this sort, in an operating environment optimised to investigate the long-term growth and development of animal cognition, one can see the first indications of auto-regulatory behaviour in a serial, exhaustive search task. Following the expression of class inclusion management and categorical boundary effects (measured as a function of the reaction times to touch successive icons in the sequence) in the sequencing of nine-item strings composed of three categories, each with three exemplars free to vary in size or colour, the five monkeys so far tested have self-imposed serial touch pattern constraints upon themselves

in a controlled and principled way, well beyond the requirements of the experimental task as given. Indeed, such self-imposed sequential ordering with the categories proved to be idiosyncratic acros monkeys, and as such has also resulted in the production of quite different RT profiles for each, according to the preferred sequence and their associated search pattern requirements (see table 4. 45). This latter finding, true only for the more salient, non-arbitrary, size-relational equivalence phase, provoked the emergence of the first variations in chunking for a nine-item sequence, deviating from the more usual three-x-three-x-three phrasing pattern.

Table 4.45 Monkey subjects showing significant latency rise-times for equivalence phases, occurring only at the category boundaries

Subject	Physical		Colour		Size	
	1st	2nd	1st	2nd	1st	2nd
Charlie	◊	x	◊	x	◊	◊
Alfie	◊	◊	x	x	◊	x +9
Mimi	◊	◊	◊	◊	◊	x
Luba	◊	◊	◊	◊	-	
Ollie	◊	◊	◊	◊	-	

◊= ($p<0.05$); x = NS; +*n*,..= other sig. ($p<0.05$) RT position (s).

Such findings offer confirmation that there is indeed a method here available for the capturing of a dynamic-interactive epistemological growth in the monkey subject, both informed by task success and informing the researchers, following the years of preparation and patient step-by-step evaluation of competences thought to underlay cognitive growth and development.

(C) [DEF] Transfer Acquisition Phases

Introduction

It had now been firmly established that the *apella* monkey was capable of demonstrating consistent seriation skill with a number of nine-item string-length sequences. Whether this species could tolerate further extension of their existing series' [ABC] icon sets in terms of breadth (say, to fifteen or twenty items) still remained unknown, but it was also important to determine whether the extent to which these high levels of performance attained might have been the result of the particular [ABC] incremental procedure used. One test of this concern would be to

investigate the extent to which the monkey might generalise their serial, exhaustive search skill performance to equally long sequences comprised of previously unseen icon sets as measured by their acquisition and serial order control performance. However, in contrast to the previous phases, for which all monkey subjects achieved successful performances with nine-item, three category classifications, [ABC] multiplex and equivalence trials, a second set of acquisition phases were given with a view to assessing the extent of any possible transfer effects which could be made evident using an entirely novel three-category core sequence given cold. If tolerant of the novel stimuli, as measured either by success and/or in the numbers of trials to criterion required in acquisition rates for the [ABC] sets of equivalent string-lengths, measures of transfer might inform us as to what it was that the monkey was learning, independent of the actual stimuli set being employed. Was the subject merely coping well with longer sequences by memorial processes primed to the use of the absolute stimulus properties of the by now familiar [ABC] icon sets ?, or was there now evidence for a more general cognitive skill developing, in that the monkey would indeed prove capable of demonstrating spontaneous classification for any set of icons divisible into three categories, in arrays of such large sizes as now traversable?

The [DEF] transfer phases 15 through 17 to be employed were as shown in Table 4.46 below and include the details of interspersed conditions and completion criteria for each phase.

Table 4.46. Acquisition Phases 15-17 [DEF]

Phase No.	Conditions	Completion criterion
15	DEF	15/20
16	DEF DDEEFF	3/4 15/20
17	DDEEFF DDDEEEFFF	3/4 15/20

All subjects participating in this experimental series started with Phase 15. Each phase required constant attention and vigilance, with consistently high levels of performance being maintained in the face of each successive phase again subsequently representing greatly increased levels of task difficulty. For each of the transfer phases, individual

monkeys were self-selecting by their continuing successful attainment of the given phase criteria as they progressed through the tasks. Phases were given in the order shown and there was no explicit tutoring of the monkey for single unique solutions for any excepting the first [DEF] core-sequence condition. For all other conditions in each of the phases 16-17, only [D..E..F..] category order was required to be sequentially maintained.

Stimuli

Each subject was assigned a new three-item category set [DEF] each comprising a novel shape and colour, both from each other and distinguishable from the stimuli used in their individual [ABC] sets (e.g. cyan star [D], yellow circle [E] & magenta hourglass [F]), controlled for stimulus-preference effects across animals. For use also in the later equivalence phases, each subjects' [DEF] stimuli pool contained three coloured, two monochrome and three sizes of each icon as before. As a general code, the following key provides examples of the nomenclature to be used in identifying string composition throughout the following Phases 15 through 22:

D= 1st category shape (e.g., cyan triangle)
E= 2nd category (e.g., green circle)
F= 3rd category (e.g., magenta hourglass)
D_C= 1st category shape free to vary in colour (e.g., red, white or blue star)
D_S= 1st category shape free to vary in size (small, medium or large star)

Phase 15

Rationale. To assess the degree to which establishment of the core sequence [ABC] and its subsequent extension might be generalisable to the control of serial order production with other categories (and not merely as a response to the nature of the incremental procedure or actual stimuli previously used) a new core three-item string [DEF] was presented for discrimination and explicit ordering. This first phase would inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli set in a consistent and orderly serial manner. Further, if the monkey was responding with a transfer of serial skill to the novel three-item set, it would be expected that each subject reach criterion in a

shorter period of time than was required to reach the equivalent stage with their [ABC[sequence.

Subjects. Three adult monkeys took part in this phase (Charlie, Alfie and Mimi), all of whom had completed twelve-item seriations and equivalence trials in which nine-item [ABC] icon sets were free to vary in size or colour.

Design and Procedure. In accordance with the general procedures used above, this single training phase condition [DEF] employed a three category icon string, presented cold without a [DE] precursor as before (cf: phases 1-3) each icon discriminable by both colour and shape. The subjects' task was to sequentially interrogate and touch its category icons [D] first, followed by [E], then finally [F], with four possible error-types (touching icon [E] or [F] first, [D] followed by [F] - a forwards error, or [D] [E] [D] - a reiteration), for which at no time was any differential feedback given for incorrect trials. The experimental criteria for success with this 'core sequence' phase was again at the 75% level, requiring a proficiency maintenance of fifteen correct out of the last (cumulative) twenty trials, after which, if successful, the session was automatically halted. To assure the establishment of this novel core sequence [DEF], the phase was consolidated by each subject being required to reach the task criterion twice before moving on to the following duplet string phase.

Results. All three monkeys completed this 'training' condition and thus showed themselves capable of reporting a second three-item, three category sequence [DEF]. The number of trials, errors and percentage errors to criteria for the condition is shown in Table 4.47.

Table 4.47 Nos. of trials, errors and % errors to criterion [DEF]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	128	76	59.3%
Mimi	753	543	72.1%
Alfie	498	361	72.5%

In comparison with their [ABC]- acquisition measures, both Charlie and Mimi required fewer trials to reach criterion for the new [DEF] condition (17% and 66% respectively), although Alfie needed 40% more trials, a

figure commensurate with those taken for him to satisfy his original [AB] condition beforehand. Of the four possible error-types, reiterative errors [DED] were rare (less than 15% overall) and the single step forwards-error [DF] accounted for some 60% of all errors except for those of Alfie. The latter monkey showed a more even spread of error-typologies for the forwards-error trials [E 1st](43%), [F 1st](25%) and [DF](31%), but reiterating on only relatively few incorrect trials (9%). Reaction time measures for the three touches were consistent across subjects and produced analysis of variance ($F_{2, 15}$=11.43, $p< 0.001$) for sequence position effects. Figure 4.15 shows the RT profile for pooled data for the criterion-run trials following acquisition. All pairwise comparisons of the RT differences for adjacent touches in the sequence were also significant ($p< 0.05$). Each monkey reported correct trials in their first session, with Charlie making as many as 25% correct in his first twenty trials.

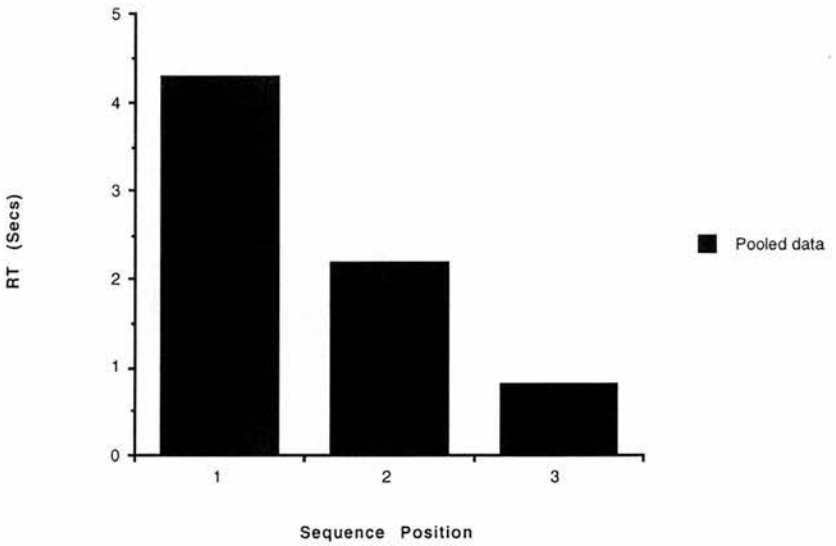


Figure 4.15 Mean RT profile for successful novel 3-item sequencing of three categories, each with a unique (colour/shape) exemplar in Cebus monkeys (N = 3, pooled)

Discussion. The successful performance of each monkey in this first [DEF]-transfer phase has shown that the choice stimuli for the initial two category shapes were readily discriminable and consistently orderable at a sustainable level over several daily sessions. It has been established that each monkey would consistently report their allocated 'core' three-item [DEF] sequence following a supervised training procedure which allowed

no flexibility of ordering on the part of the subject. Compared to the earlier [AB] to [ABC] training phase, it can be seen that the immediate, cold presentation of three novel items has resulted in a decrease in the number of trials required to reach the criterion level of performance. Indeed, even in the presence of the enlarged initial error space (from one to four error types) fewer trials were required for two subjects. For the pooled data, most of the errors recorded were of one major type, [DF] (>60%) consistent with the view that the monkey's realisation of the initial icon to be touched became quite quickly established. However, the relatively low incidence of the other error types does not necessarily suggest that the monkey was demonstrating high degrees of transfer of class ordering from the earlier training phase [ABC]. Indeed, especially for the error profile of Alfie, the findings are consistent with an associative-chaining explanation for a three-item set. The RT findings of all six monkeys, however, in which the time taken to search each subsequent icon of the set to be ordered decreased as successive items were interrogated, is suggestive of the development of a route-planning strategy to be taking place within each trial. Only with the use of somewhat larger, novel icon sets, would the issue of the presence of classificatory transfer skill be addressed in sucha way that might not allow an associative-chaining explanation.

Phase 16

Rationale. With a view to determining whether the monkey were going to be capable of the spontaneous classification of increasing numbers of a now familiar second set of icons in the array, the next phase set out to explore the possibility of serial order skill transfer to a novel six-item set. By increasing the sequence length (again by breadth) without changing the previously learned ordinal positions of the individual categories to be reported, an further item was added to each category simultaneously to produce a duplet six-item string [DEF] -> [DDEEFF]. The doubling of the set size in this phase of the [DEF] core-sequence expansion would not only offer an indication of the monkey's ability to spontaneously classify a principally orderable set. Should the monkey succeed, such success would provide good evidence for the imposition of a classificatory scheme by the subject in order to assist the serial production of a multiple-item three-class series.

Subjects. Two adult monkeys (Charlie and Mimi) took part in this phase. A third monkey subject (Alfie) was taken by alternative single-step increments towards nine-items, so as to help temper what appeared to otherwise be too larger demands being placed upon him with the increasing set sizes (and thus cognitive load).

Design and Procedure. Having successfully completed the [DEF] training phases, all subjects qualified for this, the first of two [DEF] transfer phases (with criterion) which allowed the monkey to demonstrate the ability to make larger transfer steps with the addition of one (physically identical) item to each category simultaneously, so producing duplet strings [DEF] -> [DDEEFF]. The two conditions were presented at a ratio of 1:4 for single [DEF] to duplet [DDEEFF] sequences, the latter offering error-types of twelve possible paths with the six-item set. Following every twenty duplet string exposures, maintenance levels of three out of four successful [DEF] trials were required in order that the [DDEEFF] test condition trials continued to be presented. The task criterion was set at fifteen correct trials out of the last (cumulative) twenty trials and when satisfied, lead to the termination of the session.

Results. All three monkeys completed this 'transfer' condition and successfully reached the level of performance as set by the experimental criteria, whilst maintaining 75% correct trials with the control condition. The number of trials, errors and percentage errors to criteria for the test condition is shown in Table 4.48.

Table 4.48. Nos. of trials, errors and % errors to criterion [DDEEFF]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	422	273	64.7%
Mimi	586	383	65.3%
Alfie*	184	96	52.2%

* = dddef

Error profiles reveal that of the twelve possible ways of committing error, only four were to any degree represented. 95% of errors occurred at the category boundaries [Dx] (50%), [DDx] (10%),[DDEx] (36%), the remainder being reiterative errors, only 13 trials involving first-position touching. Reaction time measures indicate significant differences between subjects,

both in absolute execution times and in profiling with respect to sequence position. Individual profiles for Charlie and Mimi are shown in figures 4.16-17, and it can seen that Mimi was taking two to three times as long to make her touching throughout the sequence. Individual RT differences at category boundaries (this time between positions three and five) were not significant for Charlie, but for Mimi there was a real effect for both boundaries ($p < 0.05$), despite the unusual direction of the first.

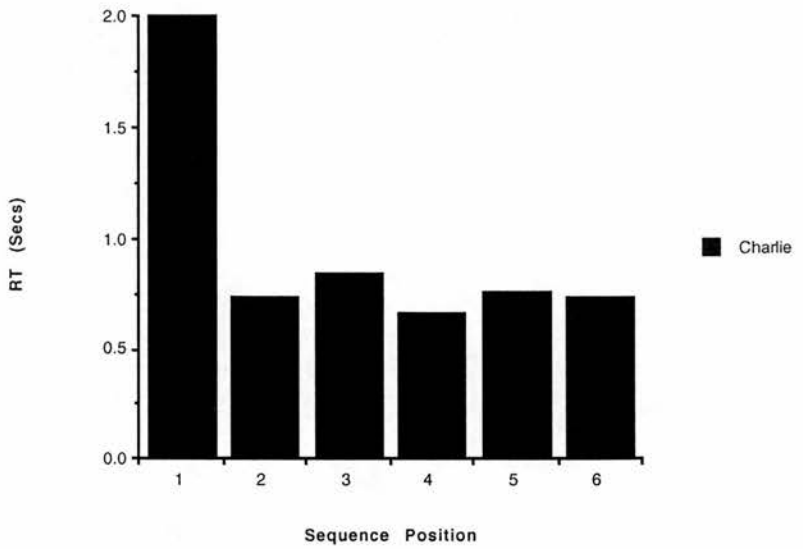


Figure 4.16 Mean RT profile for successful novel 6-item sequencing of three categories, each with two identical exemplars in a Cebus monkey (Charlie)

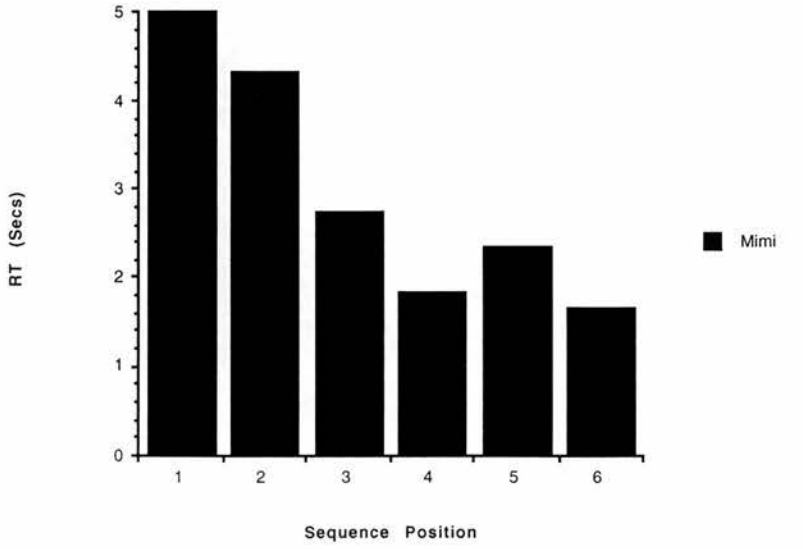


Figure 4.17 Mean RT profile for successful novel 6-item sequencing of three categories, each with two identical exemplars in a Cebus monkey (Mimi)

Discussion. Condition criteria were met by both of the monkeys in this phase, consistently reporting interspersed three- and six-item (duplet) strings. This success readily upheld the same three category markers with each maintaining their relative sequence order as presented in the previous [DEF] training phase (although again possibly at variance with their absolute ordinal positions). The absolute numbers of trials needed prior to reaching the condition criteria were roughly comparable with those taken on the [ABC] six-item (duplet) strings, although differences in error patterns were observed. First touch errors were rarely recorded at all, especially notable now that the insertion of a second exemplar for each category in the second position had further removed the option of succeeding in the task by means of categorical chaining alone. Despite the possibility of twelve error types for the test condition, error profiles revealed clear preferences for touching in accordance with the previous [DE] and [EF] adjacency sequencing once, these accounting for more than 85% of all the error trials for combined errors in test condition. Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner cannot be known due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (any forwards-error touch ended the trial). However, it was now becoming quite clear that the monkeys were quite capable of demonstrating their proficiency with a novel six-item sequence under supervised-learning conditions for which this degree of free search was permissible. Evidence for the existence and use of some strategic planning (or at least economic management) with these larger sequences come not only from an analysis of error frequency and the shift towards dominantly forward looking errors, but were also becoming evident following the development of clearly differentiable, though not highly significant, RT profiles for the new six-item set. The findings indicated that compositionality was indeed a factor in determining the nature of the RT profile for sequence position, the profile contour reflecting the nature of the categorical boundary features of the condition being interrogated. Further, much of the time taken to successfully complete each trial was used up in the period before the first touch had been made, indicative of careful inspection and forward planning prior to sequence execution. These findings provide further evidence indexical of the classificatory processes being revealed by the

monkey. If these claims were indeed to be supportable, then such phrasing effects might be expected to persist (or to become even more salient) with increasingly larger string-length and their compositional variations.

Phase 17

Rationale. Having by now satisfactorily demonstrated the ability to negotiate a novel six-item sequence composed of three categories, each with two exemplars in the last phase, it was now reasonable to give the monkey a second transfer phase, again, with the simultaneous addition of one item to each category in a single step. The monkey would thereby inform us as to its ability to develop the core [DEF] sequence so as to produce a string of sufficient length suitable for the later introduction of a depth of search,*within* a category (see figure 3.3). Furthermore, indications as to whether the subject might be able to consistently report a novel nine-item stimulus array in a systematically controlled and principled way might be obtained without this time having exposed them to other multiplexed combinations of the same icons along the way. If the benefits of classification skills are to be seen within this phase, then one might expect to observe the monkeys to be completing the phase with far fewer trials and errors being required to reach criterion levels of performance, and, more pronounced category boundary effects to occur with the longer nine-item sequences (by now difficult to be explained by a simultaneous-chaining account).

Subjects. Three adult monkeys took part in this phase (Charlie, Mimi and Alfie). Each previously having been given three- and six-item strings of the [DEF] stimuli sets and having reached criteria levels of performance.

Design and Procedure. Continuing with the extended [DEF]- core sequence for each monkey, this phase presented interspersed control trials [DDEEFF] to assist the interpretation of any discovered failure to maintain or improve performance with the novel nine-item sequence [DDDEEEFF]. As in the previous phase, each subjects' task was to sequentially interrogate and touch all of its category [D] icons first, followed by all of the [E], then finally all of the third category [F]. The two conditions were presented at a 1:4 ratio for duplet and triplet sequences respectively, employing multiple

exemplars of the same icons of categories [D], [E] and [F] as previously used. Following every twenty triplet string exposures, maintenance levels of three out of four successful [DDEEFF] trials were also required in order that the [DDDEEEFFF] condition continue to be presented. The phase was automatically completed when the triplet condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All of the monkeys completed this second 'transfer' condition and were thus eligible for continuation to the [DEF] equivalence phases. Whilst simultaneously satisfying the 75% criterion required for the control condition [DDEEFF] the mean number of trials, errors and percentage errors to criteria for the test condition is shown in Table 4.49.

Table 4.49 Nos. of trials, errors and % errors to criterion [DDDEEEFFF]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	496	297	59.9%
Mimi	96	49	51.0%
Alfie	127	67	52.7%

All subjects showed a number of correct trials in their first session and in the case of Charlie, there was to be a 30% hit rate in his second batch of twenty trials. Accounting for between 78% and 93% of the total error trials for any one subject, it was seen that the two single-step forwards-error mistakes [DDE] and [DDDEEF] were the dominant error-types despite there now being a total possibility of making any of twenty-one error types. (The error profile for the control condition [DDEEFF] was roughly comparable with those reported for it as the test condition above, with the two dominant error types maintaining their share of the total error space at levels of 85%, (31% for [DE] and 54% for [DDEF] respectively). RT data once again showed high significance for analysis of variance ($p<0.01$) for all subjects, but the significance levels of actual pairwise comparison data for particular positions (including category boundaries) was mixed across subjects. Due to this large variance, the three subject's data could not be reliably pooled and so individual subjects profiles are presented below (see figures 4.18-20). For Charlie's profile, although the reaction times do indeed increase at both category boundaries (sequence positions four and seven), these did not do so significantly. There was, however, an unusually

noticeable and significant ($p<.0.01$) rise in the search time for the last item. In the case of Mimi, again much slower than Charlie although more accurate, the RT profile reveals a significant second category boundary effect only ($p<0.005$), similar to that latterly found for her previous [ABC]-equivalence phases.

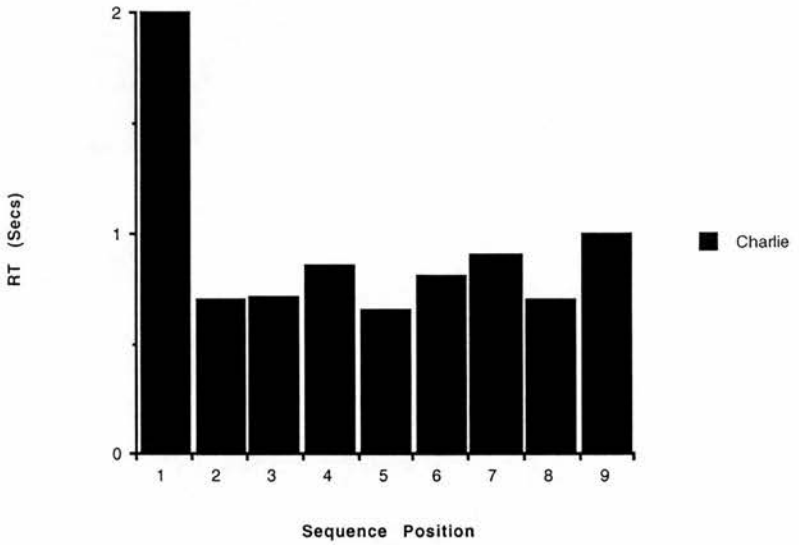


Figure 4.18 Mean RT profile for successful novel 9-item sequencing of three categories, each with three identical exemplars [DDDEEEFFF] in a Cebus monkey (Charlie)

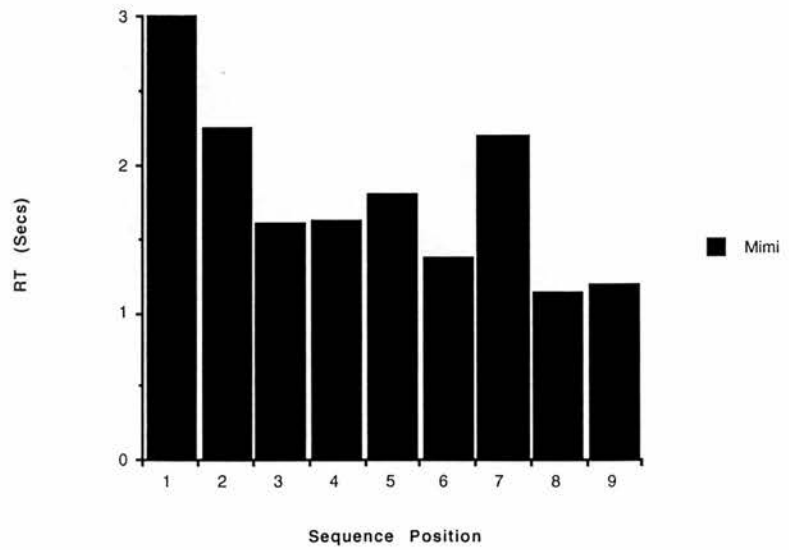


Figure 4.19 Mean RT profile for successful novel 9-item sequencing of three categories, each with three identical exemplars [DDDEEEFFF] in a Cebus monkey (Mimi)

For the third subject, Alfie, both boundary effects visible, but again, as with Charlie, only the second of these (sequence position five) showed a significant difference ($p<0.05$). The only other Bonferoni post-hoc evaluations to showed significant effects were for decreasing RT values at positions two and five.

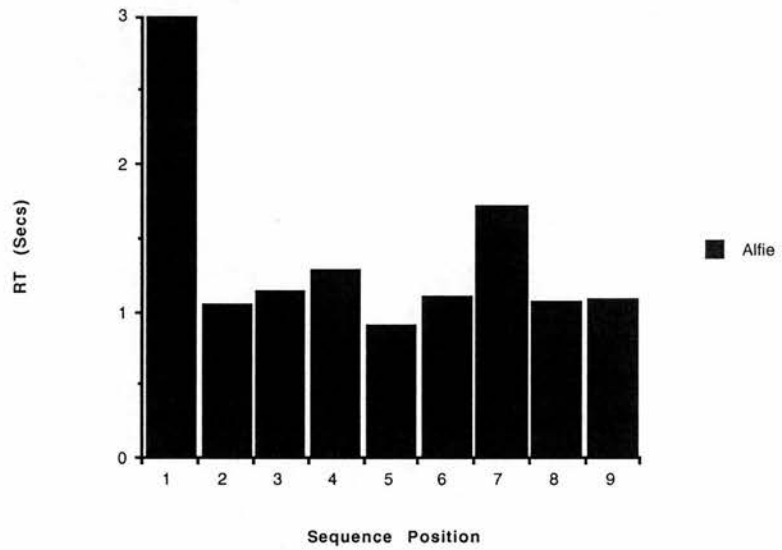


Figure 4.20 Mean RT profile for successful novel 9-item sequencing of three categories, each with three identical exemplars [DDDEEEFFF] in a Cebus monkey (Alfie)

Discussion (and interim summary of acquisition transfer phases 15-17). Although there was a large step from six to nine items in this last phase, both monkeys (and a third following a single-step incremental pathway) succeeded in reporting correct sequencing of a novel [DDDEEEFFF] nine-item string despite the much higher level of difficulty and performance required. This has shown that the first demonstrations of the abilities of the non-human primate to successfully execute a simultaneous three category, three-exemplar sequence, are quite reproducible and indeed, within the same subject can provide indicants of serial skill transfer with faster uptake of a subsequent novel icon set. When compared to the acquisition of the nine-item [ABC] set, the new [DEF] nine-item set required of each subject took only 29%, 4% and 4% the number of trials to reach criterion for Charlie, Mimi and Alfie respectively (see table 4.50).

Table 4.50. Transfer savings measures: trials & errors to criterion [AAABBBCCC] Vs [DDDEEEFFF]

Subject	No. of Trials ABC	No. of errors ABC	No. of Trials DEF	No. of errors DEF
Charlie	1686	1114	496	297
Alfie	2801	1993	127	67
Mimi	1950	1659	96	49

Another indication of their successfully exploiting the principle of the task's simple classificatory scheme was to be seen both in the fact that all monkeys succeeded in completing a correct sequence within their first session, and that there were significant patterns in their error profile for incorrect trials. Given the possibility of as many as twenty-one error types for this novel nine-item test condition, error profiles revealed clear preferences for sequential touching in accordance with target category sequencing once more, with one-step forwards-errors ([DDE] and [DDDEEF]) accounting for more than 83% of all the error trials recorded (at a rate of 52% and 31% respectively). Whether the monkey would have continued to produce a successful exhaustive search of the array in some other ordered manner still remains unknown at this stage due to the nature of the machine constraints being imposed upon the subject by the tasks' programming requirements (any forwards-error touch ended the trial). But evidence for the existence and use of some strategic planning (or at least economic management) with movement though this much larger sequence derives not only from an analysis of error frequency and measures of transfer. The overall time taken to successfully complete each trial continues to be over-represented in the initial reaction time to the first touch (IRT). There were also other consistent patterns (although with varying levels of significance) to be found in the RT data for category boundary positions (see table 4.51).

Table 4.51 Monkey subjects showing significant latency rise-times for duplet and triplet phases, occurring only at the category boundaries

Subject	DDEEFF		DDDEEEFFF	
	1st	2nd	1st	2nd
Charlie	x	x	x	x +9
Alfie	x	◊	x	◊
Mimi	x	◊	x	◊

◊= ($p<0.05$); x = NS; $+n$,..= other sig. ($p<0.05$) RT position (s).

Taken as indicators of careful inspection and forward planning prior to

sequence execution, these findings continue to support the earlier first indications of the successful management and control of the execution of nine-item sequences by the non-human primate, and indicate levels of performance consistent with the presence of classificatory processes in the *apella* monkey, at least in the supervised laboratory environment. If this were to continue as before, then such phrasing effects might be expected to persist (or to become even more salient) either with increasingly larger string-lengths or with their compositional equivalence variations as previously seen above.

(D) [DEF] Equivalence phases

Introduction

It has now been firmly established that the monkey was capable of demonstrating consistent seriation skill with sequences of at least nine-item string-lengths using extended [ABC] icon sets both physically identical and free to vary within a given category. Whether this species would tolerate further extension of their novel series' [DEF] icon sets in terms of breadth (say, to twelve or fifteen items) is currently not known, but there was little to suggest that they were anywhere close to reaching of an upper limit of performance in the case of nine items. Indeed, all monkeys completed their [DEF] nine-item sets in a much shorter time than was taken to satisfy the criteria levels of performance for each's own [ABC] nine-item sequence, suggesting a much reduced level of effort now being required to solve seriation management with these larger set sizes. However, having seen all monkeys now achieve successful levels of performance with a second nine-item, three category classification level, rather than continuing to increase the number of exemplars in each category (resulting in a search problem of its own) we continued to investigate the possibility of any depth of search transfer that the monkey might be able to report. One way of decreasing the memorial load being placed upon a subject working with very long sequences, would be for them to attend to the presence of any distinguishing features salient to the individual items to be seriated, which could then be used in some hierarchically organised way. The problem would thereby be solved by selectively reducing the larger search space into more manageable chunks as had begun to emerge in the [ABC]- equivalence phases. Following the completion of their new nine-item string lengths

[DDDEEEFFF] they were now to be exposed to two forms of physical variation, again to be presented as equivalence phases. Although it might now appear quite likely that the monkeys might have at this time proceeded to spontaneously report principled ordering of icon sets with categorical exemplars varying in colour and/or size under free-search conditions without much difficulty, at this stage of our exploration of transfer function characteristics, it remained important to first establish the degree of the monkeys' tolerance to such variation in the appearance of their new icon sets under supervised conditions. In order therefore to ascertain whether the feature variations to be used for each subject in the future were to be manageable by the monkey, a probe set of equivalence phases was undertaken by each. At this stage there would still be no requirement for the subject to explicitly order the new nine-item sets according to any particular size or colour code (although the subject could self-impose such a constraint), the aim here was once again to simply see whether the monkey would maintain its existing levels of performance with a nine-item, three category set whilst at the same time tolerate feature variation within each category. Such steps are important, because without the knowledge of the monkey's ability to distinguish them, let alone be able to view them as belonging to a class-equivalent set, one would not be in a position to readily interpret their performance with an explicitly hierarchical search task, especially in the face of their producing large amounts of failure. The [DEF] equivalence phases 18 through 20 to be employed were as shown in Table 4.52 below and includes the details of interspersed conditions and completion criteria for each.

Table 4.52. Equivalence Phases 18-20 [DEF]

Phase No.	Condition	Completion criterion
18	DDDEEEFFF - monochrome	15/20
19	DDDEEEFFF- monochrome	3/4
(colour)	$D_CD_CD_CE_CE_CE_CF_CF_CF_C$	15/20
20	DDDEEEFFF- monochrome	3/4
(size)	$D_SD_SD_SE_SE_SE_SF_SF_SF_S$	15/20

Two subjects participating in this experimental series started with Phase 18, a third, Charlie, (who had been already working with monochrome contours on a sideline selective loading project for a while) started at

Phase 19. Each phase required an extremely high level of vigilance and consistently high levels of performance to be maintained (with thirty possible error types being possessed by each condition). Notwithstanding icon differentiation problems, each condition was in principle, an equally difficult task. Phases 19 and 20 were given in randomised order between monkeys, some receiving 19 before 20, and others the converse. These two phases presented three exemplar, three category nine-item sequences to be exhaustively interrogated as before, but presented icons that could again each be uniquely identified within the array; either by colour [$D_CD_CD_CE_CE_CE_CF_CF_CF_C$] or by size [$D_SD_SD_SE_SE_SE_SF_SF_SF_S$]. The subject was required only to report the sequence by category as before (i.e., touch all [D_X] before all [E_X] before all [F_X]) and at the same time to maintain levels of three out of four successful (physically equivalent) control trials. These latter equivalence phases were to complete the acquisition and equivalence transfer trials for the three-category [DEF] series.

Phase 18

Rationale. In order that the ensuing equivalence phases be consistently interpretable for both colour and size variations, it was important that the baseline nine-item set be devoid of colour cues from the outset. In order to achieve this, the existing [DEF] categories' shapes as previously used in the nine-item sequence [DDDEEEFFF] were maintained, but following its colour feature subtraction, was then presented as monochrome contoured (familiar) shapes for base-line comparison (e.g., white circle outline on a black background). The [DDDEEEFFF]-monochrome contour stimuli set was also to form the base-line condition for the equivalence phases which were to follow. The subject's performance on this phase would also inform us as to whether the previous transfer characteristics of successful seriation might be due merely to familiarity effects related to the particular stimuli employed, or indeed to the supervisory nature of the incremental technique used to produce these long sequences in the first place. Bearing in mind that although the subjects were now two to three years into the experimental series, his would be only the second time that any monkey was to view a novel nine-item array containing previously unseen icons for a serial, exhaustive search task.

Subjects. Two adult monkeys took part in this phase, Mimi and Alfie (Charlie had meanwhile captured experience with multiple exemplar arrays with monochrome contour [DEF] sets whilst working with a short selective loading experiment).

Design and Procedure. This new nine-item [DDDEEEFFF]-monochrome contour base-line condition was given without a confounding control, providing a singly consolidated criterion for three familiar categories (at least by shape). As for all of the previous phases, each subjects' task was to sequentially interrogate and touch all of its category [D] icons first, followed by all of the [E], then finally all of the third category [F]. For any given trial, the three icons of each category were only distinguishable by their spatial location on the touch screen. This single condition presented error-types of twenty-one possible incorrect paths with its nine-item string length comprising three categories, each possessing three physically equivalent monochromatic icons The phase was continued until performance levels met the usual task criterion of fifteen correct out of the last (cumulative) twenty trials.

Results. Both monkeys achieved the experimental criterion level for the new physical equivalence condition, and did so in far fewer trials than that used to have successfully completed their previous [AAABBBCCC-monochrome contour] nine-item sequences. Their mean number of trials, errors and percentage errors to criteria for the [DDDEEEFFF-monochrome contour] condition is shown in Table 4.53 below.

Table 4.53 Nos. of trials, errors and % errors to criterion [DDDEEEFFF Monochrome Contours]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	-	-	-
Mimi	140	53	37.8%
Alfie	104	61	58.6%

Although there were twenty-one possible error-types for this condition, in the case of both subjects, more than 88% of all errors were to be accounted for by two forwards-error types. The possible errors which could have occurred in sequence positions one, four, seven and eight (fifteen in total) were hardly represented at all in the monkey's error profile, and zero values were often reported. Both of the dominant error-

types had in common that they represented one-step, feed-forward errors at the category boundaries [DDE] (43%) and [DDDEEF] (39%) in the third and sixth positions. Category boundary effects were also in evidence following analysis of individual reaction time profiles which show Mimi reporting increased time to be taken at both the [AB] and [BC] boundaries, and Alfie to be doing so for the latter boundary only. For individual animal profiles, only Alfie's second category boundary RT difference was significant ($p<0.05$) despite the appearance of his profile (see figure 4.21).

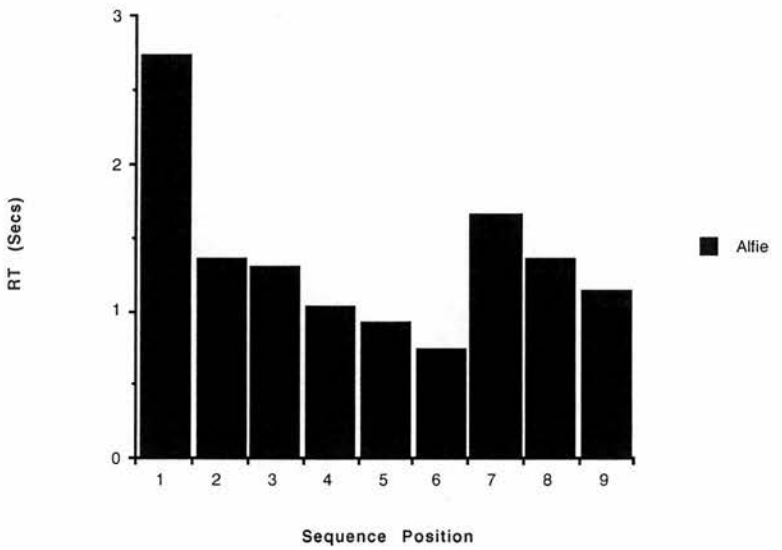


Figure 4.21 Mean RT profile for successful novel 9-item sequencing of three categories, each with physically equivalent exemplars in a Cebus monkey (Alfie)

Discussion. This equivalence transfer phase has revealed once more, an ability in the monkey to successfully seriate a nine-item set composed of novel icons, the absolute properties of which they had never before seen. Although preservative of outline shape, the stimuli of the array were similar to those of each monkey's previous sets, only having undergone cue subtractivity, in that each of the three category icons now had both colour and shading densities removed. Despite this loss of stimulus property, all subjects showed fast uptake and sensitivity to the remaining icon features, and in comparison to the original [AAABBBCCC-monochrome contour] acquisition data, the new arrays were successfully manipulated with between 80 to 90% fewer error trials being demonstrated by one monkey before reaching the phase criterion for completion.

Likewise, the total number of trials to criterion were also markedly less (in the case of one monkey, from previously more than a thousand trials to this time just over a hundred) with all subjects requiring fewer than 80% of the number of trials in order to have met their criterion-run than did they for the earlier coloured-icon seriations. The actual percentage of incorrect trials to criterion in this phase also fell by at least 10% for each animal. Error types were very informative with most of the cell counts remaining at zero, and almost all of the errors committed becoming exclusively accounted for by a one-step forwards-error seen at either one of the category boundaries. Of these, more were occurring at the later [DE] boundary, which was also indicated to be a significant ordinal position for successful trials. As indicated by the individual RT profiles, the inspection time for the first touch of each new category icon in the sequence remained significantly longer than were the times for the intra-category touches. Furthermore, the absolute time for the beginning of each category continues to become reduced, suggesting that the subject is now able to 'chunk' the larger array into smaller bites relatively quickly, and perhaps to make better predictions concerning the intra-class touches (based upon location identification only) once the salient common icon feature has been identified. The later class boundary would appear to remain a longer pause, a time during which possibly a fresh scan of the array is being made and a renewed motor pattern is being put into action. Clearly indicative of good acquisition skill transfer and clear generalisation of class inclusion ability, this second equivalence test of a monkey's strategic competence with serial order control of large lists continues to suggest that the monkeys are now informing us that they were now quite well task-informed.

Phase 19

Rationale. Having now demonstrated a second set of successes with a novel nine-item sequence with physical equivalence characteristics, it was now possible to explore the degree to which the monkey might show evidence of transfer with feature variation within a class, whilst maintaining categorical seriation. In order that the animal might inform us of their abilities to spontaneously classify in a more free-search situation, it was first necessary for us to determine whether they might report class inclusion for 'known' category icons which might vary along some

experimentally manipulated dimension (for example, colour) under supervised learning conditions.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This second non-physical-equivalence phase presented another completely novel looking three category, three exemplar, nine-item sequence to be exhaustively interrogated as before, but now, again comprised of icons that could each be uniquely identified within the array. Free to vary within each category by colour [$D_cD_cD_cE_cE_cE_cF_cF_cF_c$], the subject was nonetheless only required to report the sequence by category as before (i.e., touch all [D_c] before all [E_c] before all [F_c]). Although it was necessary for the monkey here to observe the equivalence of particular icons for the purposes of class inclusion, it was not necessary to touch particular icons in a fixed ordinal position within a category (although they could self-impose such a constraint upon themselves if they so wished). Interspersed with these trials the subject was also required to maintain control levels of three out of four successful trials of a [DDDEEEFFF-monochrome contour] control condition in order that the colour-equivalence test condition continue to be presented. The phase was completed when the equivalence condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this 'colour equivalence' task whilst maintaining the 75% criterion required for the control condition for the duration of the phase. The mean numbers of trials, errors and percentage errors to criteria for the test condition were as shown in Table 4.54.

Table 4.54 Nos. of trials, errors and % errors to criterion [$D_cD_cD_cE_cE_cE_cF_cF_cF_c$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	626	420	67.1%
Mimi	805	498	61.8%
Alfie	484	271	56.0%

Although there were twenty-one possible error-types for this condition, for all subjects, more than 82% of all errors were accountable for by two

forwards-error types. The pattern of errors was similar to that found in the last phase, with both of the dominant error-types representing one-step, feed-forward errors at the category boundaries [$D_cD_cE_c$], and [$D_cD_cD_cE_cE_cF_c$], in the third and sixth positions respectively. Category boundary effects were also in evidence following analysis of variance for each subject's reaction time profiles, but due to unequal variance could not be pooled. The data for Alfie demonstrated category boundary effects, with increasing times to touch at both the [D_cE_c] and [E_cF_c] boundaries, but only the latter was significant ($p< 0.05$). Mimi showed both fall at the first and rise at the second, but neither were significant at below the 5% level. Charlie, however, was to show the most unusual RT profile in that his showed significant pause times at unprecedented positions five and six in the sequence, followed by a significant drop in RT at the last category boudary (all $p<0.05$, see fig. 4.22). Some evidence for spontaneous ordering effects also emerged for two of the subjects, and for Charlie, the two white outline contour icons of the first two categories were most often touched in positions one and six, and there was some tendency for him to complete his sequence by starting with the white, and ending with the red coloured icons of the last category.

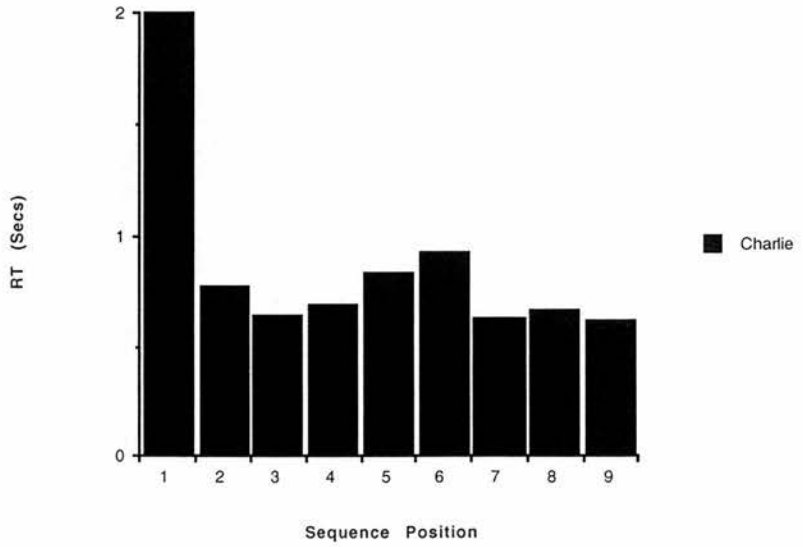


Figure 4.22 Mean RT profile for successful novel 9-item sequencing of three categories, each with colour equivalent exemplars in a Cebus monkey (Charlie)

Mimi was most likely to commence each category with the outline contour shape, but was equally likely to touch either of the remainder in the

second or third places. Alfie showed no preferences at all for explicit ordering and behaved as if true equivalence was an efficient enough strategy to endure the sequence to its end on each trial.

Discussion. This phase has provided another demonstration of the non-human primate successfully seriating a nine-item set composed of three categories, within each of which are three exemplars free to vary in their coloration characteristics. Although preservative of outline shape, the stimuli of the array were similar to those of each monkey's previous phase, but this time having undergone cue additivity, in that each of the three category's icons now had both colour and shading density variations. All subjects showed even faster uptakes and sensitivity to the icon features of their stimulus arrays, and in contrast to the previous novel nine-item [$A_cA_cA_cB_cB_cB_cC_cC_cC_c$] acquisition data, the new [DEF] arrays were successfully manipulated with far fewer trials (60% and 76%) for two of the subjects. Charlie appeared to take significantly longer (requiring some 30% more) although still requiring much fewer trials than that needed to complete their original [AAABBBCCC] nine-item set. Error types were very similar to those seen with the physical-equivalent string in the previous phases. It is possible that the changes in RT profile related to the emergent use of spontaneous ordering of the sets being interrogated. The prominent peaks in Charlie's profile for this phase were in the first and sixth positions, and it was here that the two white outline contour icons were mostly touched (the first being the old [D] and the sixth being the old [E] stimuli respectively) although the other icons in those categories were free to vary in order. There was also some tendency for Charlie to complete his sequence by starting with the white, and ending with the red coloured icons of the last category. It is possible that this new RT profiling effect reflected his explicit search for these particular intra-category icons between the fifth and sixth touches. Overall, a good acquisition and serial order skill transfer, plus clear manipulation of early class inclusion abilities were evident from the outset of the phase, in which every monkey saw at least one correct test trial in their twenty of the first session.

Phase 20

Rationale. Having seen a second group of successes with a novel nine-item sequence with colour equivalence characteristics, it was now pertinent to explore whether the monkey would tolerate further feature variation within the same classes of the same categorical seriation task. Again, in order that the animal might inform us of their abilities to spontaneously classify in a more free-search situation, it was first necessary for us to determine whether they might continue to report class inclusion for their 'known' category icons free to vary, this time, in size, under supervised learning conditions.

Subjects. Three adult monkeys took part in this phase. One of the females (Mimi) had recently and unexpectedly given birth, but continued to present and join the daily experimental sessions (bringing her infant along to work with her) although she often did not complete her sessions.

Design and Procedure. This second of the non-physical-equivalence [DEF] phases presented another novel three category, three exemplar, nine-item sequence to be exhaustively interrogated as before, but on this occasion, and for the first time, comprised icons that could again each be uniquely identified within the array. Free to vary within each category, this time by size [$D_S D_S D_S E_S E_S E_S F_S F_S F_S$] the subject was required to report the sequence as before (i.e., touch all [D_C] before all [E_C] before all [F_C]). So although there was need for the monkey to observe the equivalence of particular icons for the purposes of class inclusion, for this phase, it was not necessary to touch particular icons in a fixed ordinal position within a category (although they could self-impose such a constraint upon themselves if they so wished). Interspersed with these trials the subject was also required to maintain control levels of three out of four successful trials of a [DDDEEEFFF-monochrome contour] control condition in order that the size-equivalence test condition continue to be presented. The phase was completed when the equivalence condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Only two of the three monkeys achieved the experimental criterion level for the new size-equivalence condition, but both required more trials than did they use to successfully completed their original

nine-item sequences. All three subjects upheld the 75% criterion maintenance performance for the control condition throughout the duration of the phase. The mean number of trials, errors, and percentage errors to criteria for the size-equivalence condition [$D_S D_S D_S E_S E_S E_S F_S F_S F_S$] is shown in Table 4.54 below.

Table 4.54 Nos. of trials, errors and % errors to criterion [$D_S D_S D_S E_S E_S E_S F_S F_S F_S$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	965	606	62.8%
Mimi	1475	1031	69.9%
Alfie	not completed	-	-

Although there were twenty-one possible error-types for this condition, for all subjects, more than 77% of all errors could be accounted for by two forwards-error types. Both of the dominant error-types were single-step, feed-forwards errors at the category boundaries [$A_S A_S B_S$] (37%) and [$A_S A_S A_S B_S B_S C_S$] (40%) in the third and sixth positions. Category boundary effects were also in evidence (in different directions), but again data could not be pooled due to the unequal variance shown between the monkey's RT data sets. Analysis of individual monkey reaction time profiles show Mimi (see fig. 4.23) to have had significantly increased touch times only at the [AB] and [BC] boundaries ($p < 0.01$) and at the sixth position ($p<0.05$).

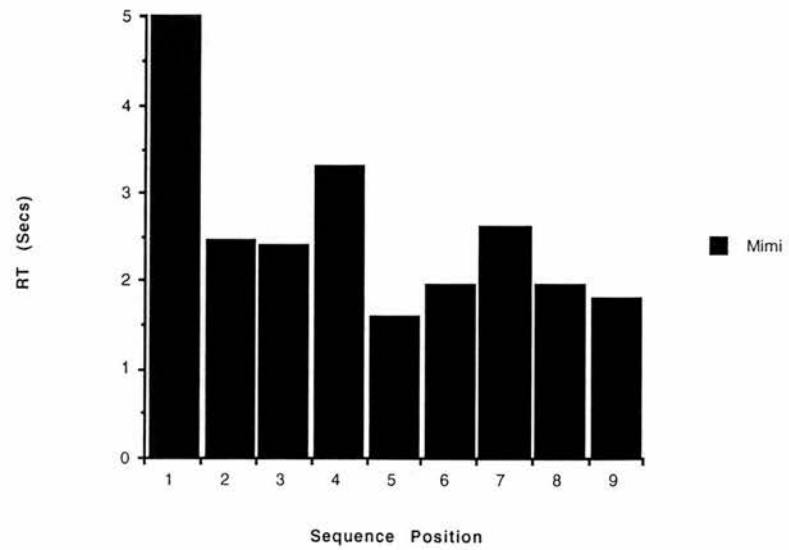


Figure 4.23 Mean RT profile for successful novel 9-item sequencing of three categories, each with size equivalent exemplars in a Cebus monkey (Mimi)

For Charlie, there were only non-significant increases at each category boundary with the addition of a third (but also non-significant) effect at the sixth position. Analyses of the actual touch sequences for correct trials in the final criterion-run, revealed some effects of idiosyncratic icon ordering tendencies. Although specific transitional probabilities were not calculated, it is noteworthy here that for Charlie, five out of his last ten touch sequences began with a preferred choice of the first icon of each category being the middle-size icon (the control [DDDEEEFFF-monochrome contour] condition icons). Although there was a less reliable tendency to end each category by touching the largest of its icons, this self-imposed constraint did not result in the stable production of a serial ordering by either increasing or decreasing sizes as he had shown before. In the case of Mimi, there were indications of a tendency to relax into an ascending monotonic size series for her last category (40%) and would otherwise touch the medium-sized icon first followed by the largest, then the smallest (30%). This latter size-ordering preference (medium-large-small) was consistently reported for the first category in five out of her last ten correct trials, and for the another three, only the first two had been reversed. Mimi's only patterned response in production of the second category was to place the small icon in the sixth ordinal position (60%). The only tendency towards patterned sequence touching from Alfie's performance on size equivalence, was that he did seem to show a preference for starting each category of the sequence with its largest icon.

Discussion (and interim summary of acquisition transfer phases 18-20). This second size-equivalence phase has provided a further demonstration of non-human primates successfully seriating a nine-item set composed of three categories, within each of which are three exemplars free to vary in their size characteristics. Although preservative of outline shape, the stimuli of the array were similar to those presented in each monkey's previous phase, but underwent cue additivity, in that each of the three category's icons were individually discriminable from each other by size. Subjects showed variation in their degrees of initial uptake and sensitivity to the icon features of their stimulus arrays, and in contrast to the previous nine-item [$A_SA_SA_SB_SB_SB_SC_SC_SC_S$] acquisition data, the new array, although successfully manipulated by all monkeys, were completed much

more slowly in terms of trials and errors to criterion (see table 4.55-56).

Table 4.55. Transfer measures: trials & errors to criterion - colour-equivalence phases [AAABBBCCC] Vs [DDDEEEFFF]

Subject	No. of Trials ABC	No. of errors ABC	No. of Trials DEF	No. of errors DEF
Charlie	438	271	626	420
Mimi	1346	938	805	498
Alfie	636	468	484	271

Table 4.56. Transfer measures: trials & errors to criterion - size-equivalence phases [AAABBBCCC] Vs [DDDEEEFFF]

Subject	No. of Trials ABC	No. of errors ABC	No. of Trials DEF	No. of errors DEF
Charlie	470	302	965	606
Mimi	476	256	1475	1031
Alfie	1361	808	-	-

Although both subjects showed early success (Charlie on trial one and Mimi by trial ten), the abnormally high number of trials (and percentage error for Mimi) to criteria might be explained by the post-natal circumstances of the working environment. Mimi was to show much task-attentional variation during her contribution to this phase (despite her persistently presenting for work) and because both Charlie and Mimi (plus infant) worked side-by-side as a social group, there was quite a lot of distraction, calling and communication ongoing between the two during the experimental sessions. Error types were very similar to those seen with the other equivalence conditions in Phases 18 and 19. Good acquisition skill transfer and clear manipulation of early class inclusion abilities were evident from the outset of the phase, in which, for example, Charlie saw 20% correct test trials during his first session. But perhaps the most noticeable finding of this phase was that both of the monkeys (and in particular, Mimi) were seen to be self-imposing an order of completion for items *within* a category, so providing a solution to the seriation problem that was indicative of spontaneous serial order control within a task not requiring such explicit ordering. There was no consistent evidence for the demonstration of spontaneous, monotonic size seriation across the entire sequence, but Mimi preferred to execute the last category by touching with a set of three icons of increasing size. For her other two categories, a more variable ordering pattern was seen but a clear preference for starting the sequence was found for the first category.

These transfer-task findings provide further evidence for spontaneous ordering effects in the monkey in response to task demands of increasing levels of difficulty. The control and maintenance of classification with the production of nine-item, three category sequences had now been shown to be repeatable with a novel stimuli set (and without using the single-stepwise incremental procedure) using the same, now more experienced and task-informed subjects.

Some three years into the program and with six monkeys content to work in the laboratory on a daily basis, it has now been clearly demonstrated that every *apella* subject tested to date will successfully produce extended core [ABC] and [DEF] sequences to at least a string length of nine-items. What all of these longer sequences had in common, however, was that they only required of the subject that they exhaustively touch each exemplar of a category before moving on to the next, in a fixed three-category order. Each category's three exemplars might shown physically equivalence or could have been free to vary in colour or size characteristics, but at all times had so far only comprised three basic categorical classes.

(D) Free search [ABCDEF] Acquisition Phase

Introduction. Now that in a position of having demonstrated the monkey's possession of two core-sequences, it was reasonable for us to present them with an extended sequence comprising six items composed of six (familiar) individual categories as determined by their already known discriminable shapes alone. This alternative method of extending the task-space introduced a new level of difficulty (whilst simultaneously affording the experienced subject some method of using their existing strategies for coping with the new task's subcomponents) by combining the [ABC] and [DEF] sequences to provide a novel six-item [ABCDEF] set. The following group of phases (see Table 4.57) were introduced to the two lead adult monkeys as probe trials: the first offered the monkey a chance to demonstrate maintenance of both [ABC] and [DEF] independently, followed by a free search six-item set [ABCDEF] which allowed spontaneous serial ordering without the constraint of any order requirement for the first time.

Table 4.57 Combination Phases 21-22 [ABC+DEF]

Phase No.	Condition	Completion criterion
21	ABC DEF	15/20 15/20
22	ABCDEF	---

Such a novel task and stimulus array would allow us to see whether, and to what extent, the monkey might readily transfer their previous production of these components in their solving of the new six-item, six category sequence.

Stimuli. The [ABCDEF] stimuli set to be used comprised a combination of the monochrome white outline contours for the [ABC] and [DEF] cons as used by each individual monkey subject. Shape was thereby preserved and would serve as the basis for known individually discriminable stimuli-sets for each subject.

Subjects. Requiring the previous acquisition of at least two separate three-item, three-category stimuli sets, there were currently only two candidate monkeys (Charlie and Mimi) from our subject pool who, at the time, qualified for inclusion in this phase.

Phase 21

Rationale. In order to assure that any failure to produce the novel six-item, six category would be interpretable, it was important to first give each monkey some refresher trials with both [ABC] and [DEF] triplets alone, especially as the subjects had not seen the former icon set for at least some six months. It was necessary to offer these in equal numbers, until performance with each was independently at the same criterion level.

Subjects. Two adult monkeys took part in this phase.

Design and Procedure. This initial consolidation phase of the combined [ABCDEF]-phase series presented two conditions [ABC] and [DEF], each on alternate trials for the duration of the session, to be ordered as for their

counterpart triplet-category acquisition Phases 3 and 15 respectively. The subjects were required to simultaneously complete each three-item string to a criterion of fifteen correct out of the last (cumulative) one twenty trials, after which time the session would automatically be terminated.

Results.

Both monkeys completed the two triplet [ABC] and [DEF] sequences within their first session and thus qualified for exposure to the novel six-item, free search phase. The mean number of trials, errors and percentage errors to criteria for the equivalence conditions are shown in Tables 4.58-4.59.

Table 4.58 Nos. of trials, errors and % errors to criterion [ABC]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	13	0	0.0%
Mimi	20	5	25.0%

Table 4.59 Nos. of trials, errors and % errors to criterion [DEF]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	15	0	0.0%
Mimi	17	2	12.0%

100% correct scoring was reported by Charlie for both conditions and Mimi was to demonstrate 75% and 88% correct trials for [ABC] and [DEF] respectively, each monkey satisfying the phase criteria within their first session. All of the error trials recorded were of the same one-step forward typology (that was, [AC] or [DF] for the two conditions) with the other first touch and reiterative errors accounting for none of the incorrect trials at all. Despite a unique touch-sequence solution being required for each test condition, amongst Charlie's successful trials, two touch sequences, one [BAC] and one [DFE], were registered as correct (see discussion below). Reaction time profiles for either monkey in each condition were similar both to each other, and to their earlier exposures during training (see results of Phases 3 and 15 above), with Mimi being somewhat slower (and perhaps more accurate) in her screen-touching than her partner.

Discussion. The immediately successful performance of each monkey for these familiar three-item, three category sequences has demonstrated

clear memory for specific sequential class ordering irrespective of the number of items to be searched. This is especially true of the [ABC] sets which were last displayed as part of each monkey's nine-item size-equivalence phase (see phase 14 above) some six months beforehand. It has been here reestablished that each monkey would consistently report their interspersed 'core' three-item [ABC] and [DEF] sequences following a supervised training procedure which allowed no flexibility of ordering on the part of the subject,... or so it was thought at the time. A most important serendipitous discovery occurred at the time of analysing the actual touch sequences made to the screen. Two of Charlie's touch sequences, [BAC] and [DFE] (both being 'normally' one-step forwards-errors) had been registered as correct, only later realised and determined to have been the result of an experimental condition programming error. Charlie's Phase 21 was actually being run as a free search condition, with only reiterative touches of all possible errors causing the trial to end, following successful exhaustive search of all icons of the array. It is therefore interesting to note that rather than imposing a novel solution, or varying his touch pattern at all (even in the face of successful deviations), Charlie continued to touch [A] before [B] before [C] (and likewise for the [DEF] set) despite there being no machine requirement to do so. Whether this be taken as indicating perseverance in the face of success (compatible with a simultaneous-chaining account) as providing evidence for an on-line self-regulating process is contentious for such a small sequence, but more light was to be shed on this issue following the results obtained in the subsequent six-item, six category phase condition.

Phase 22

Rationale. Having now satisfactorily reconfirmed both confidence and competence with their 'core' three-item sequences (each containing three different categorical class markers by shape), the monkeys had now qualified for a further and final increase in sequence length. There had still been no indications fourthcoming as to what the upper limits of string-length for the demonstration of serial order control might be for the monkey, and so a final free-choice phase comprising a six-item, six category set was to be presented. Offered cold for the purposes of generalisation/transfer evaluation, this phase would further inform us as to whether the monkey would be able to continue differentiating between

the components of such a large multi-category bound stimulus array, and indeed be capable of executing a six-item sequence containing more than three categories at all, let alone to demonstrate the ordering of such in a systematically controlled or principled way.

Subjects. Two adult monkeys took part in this phase (Charlie and Mimi).

Design and Procedure. Extending each monkey's 'core' [ABC] and [DEF] sequences further again, this single condition probe phase introduced the monkey to a six-item, six-category set for the first time, by adding them together simultaneously in the array, and delivering them cold. The new sequence was composed of six categories, each with one exemplar [ABCDEF]. Each subjects' task as planned was to sequentially interrogate and touch either sequence [ABC] or [DEF] first, followed by all of the [DEF] or [ABC] second, dependant upon the first three touched, in accordance with the rationale above. What actually ensued (and was deliberate for Mimi's trials, conducted some three months later) was that the test condition was once more delivered as a free search task, again to have important consequences for our evaluation of the monkey's self-regulatory and self-constraining search abilities. In the experimental situation, only reiterative errors stopped a given trial, and all of the 'normal' forward-error typologies were recognised as 'correct' as long as exhaustive search requirement was satisfied, in effect affording the monkey the opportunity to freely search a combined [ADCDEF] six-item string without any specific path constraint (other than to disallow reiteration). There was no experimental criterion for success with this probe phase, and it was to be presented for two consecutive sessions according to the subjects disposition to this task of much greater complexity, each session presenting a maximum of fifty trials per session.

Results. Both of the participating monkeys achieved a high degree of instant success with this new six-item, six-category icon array, each subject performing correct trials from the beginning of their first session. The mean number of trials, errors and percentage errors to criteria for the equivalence conditions are shown in Table 4.60 below.

Table 4.60. Nos. of trials, errors and % errors: [ABCDEF]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	22	2	9.0%
Mimi	36	9	25.0%

All nine errors of Mimi were reiterative errors in the last (sixth) position and were second touches to either the [C] or [F] icons, Charlie's were similarly reiterative, one in the third position [ADA] and one in the last position back to [C]. Reaction time data was unusual, but consistent across monkeys although Mimi was much slower in all position touches as before. RT profiles for each of the subjects are shown in figure 4.24 below.

Unusually, the marked phrasing here for both subjects was seen to occur at the fifth position only. The increased time taken to touch the fifth icon was the only significantly raised latency difference for either of the subjects ($p < 0.01$) and appeared to be related to the order in which the icons chosen were touched. Each monkey demonstrated a clearly consistent and partially idiosyncratic preference for their choice of path through the six-item space, and had in common that they attempted to maintain the original serial positions of the individual components of each core sequence at the expense of keeping the core sequence' independantly ordered in themselves.

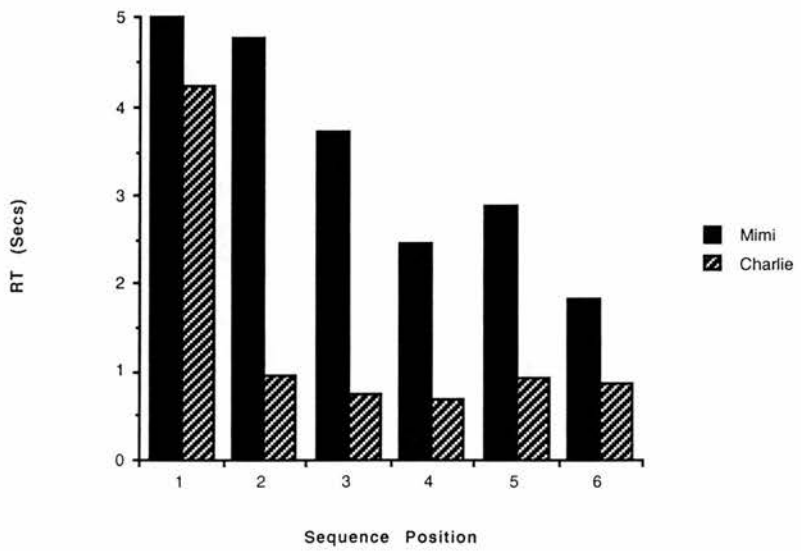


Figure 4.24 Mean RT profile for successful novel 6-item sequencing of six categories, each with one exemplar, in two Cebus monkeys (Mimi and Charlie)

Rather than offer a sequence [ABCDEF] of [DEFABC], each subject actually produced a clear preference for executing slightly different versions of collapsed pairs: [DA EB CF] for Mimi, and [DA BE CF] for Charlie. Actual touch sequences of these particular orders occurred in seven each of Mimi's and Charlie's last 10 successful trials (720 correct pathways could actually have been executed !). For the first two touches, 74% and 65% each monkey's total number of sequences started in a way suggestive of their intent to collapse their seriations by a principled first-things-first, second-things-second (and so on). On no occasion throughout the length of the phase did either subject attempt to start a sequence with either [ABC] or [DEF] triplet icon touch sequences.

Discussion. This final acquisition probe phase has provided the first demonstration with the non-human primate for successful seriations of six-item sets composed of six identifiable categories, each of which are discriminable by their shape characteristics alone. Although preservative of the outline shapes belonging to each monkey's previous phases, subjects showed immediate uptake and sensitivity to the icon features of their stimulus arrays, and successfully manipulated the novel array to produce six-item sequences which were correct from the outset. In this 'free-search' condition, error types were restricted to the making of reiterative touching due to the constraints put upon the subject by the machine programming requirements. Good acquisition skill transfer and clear manipulation of early class inclusion, and class decompositionality, were evident right from the beginning of the phase, in which each monkey saw 91% and 75% correct test trials respectively amongst their first forty trials. But perhaps the most noticeable finding of this phase was that the monkeys were seen to be self-imposing an order of completion for items, not this time *within* each 'core' sequence [ABC] or [DEF] for combined sequencing, but in collapsing icons by position *across* 'core' sequences (e.g., [AD BE CF]) so again providing a solution to the seriation problem using a spontaneous self-regulated serial order control device for a task which did not require such explicit ordering in order for the task to be solved. Although this particular icon string in the array could not have produced anything other than an arbitrary string in response (there were no other dimensions by which to rank them or sequentially determine their relationship, such as might be the case for stimuli sets

with differing size features). This finding, in contrast to those of the previous six-item conditions (see results of Phase 6) condition, was also to be reflected in the production of somewhat unusual reaction time profiles for each individual monkey. Because neither monkey chose to preserve their existing [ABC] and [DEF] 'core' sequences as a salient 'chunk', there was no increase in the times taken to touch the first icon of the second 'chunk' (as was indeed shown in a short pilot study conducted in the same laboratory with human adults). What the monkeys consistently demonstrated, right from the start of the phase, was a preferred collapsing of these two 'core' sequence's components such that a 'first-things-first', 'second-things-second', and 'third-things-third' principle emerged. The RT profiles of both subjects reflect that during the initial search time, the first four items to be identified and touched were processed as a four-item 'chunk' followed by a reflection and search for the remaining 'last-things-last' icons before the sequence's final completion. The first item of the next category could become part of that initial search now that a reduction in operational memory has been effected. What might be suggested by these highly robust and significant RT profile differences for sequence position with variable compositionality, was that perhaps there were beginnings of a case being made for these emergent phasing effects reflecting ongoing process of some kind of on-line management and control of serial order production, in the face of the increasing set-sizes to be exhaustively searched and reported. What the monkey mow appeared to be doing, was to phrase the sets into 'chunks' (either as groups of two, three or four items) as an on-line memory-reducing device, which was dependant for its detail, upon both the actual sequence length and the compositionality of the set currently under review, on-line.

Summary of results for acquisition and transfer phases.

This second implementation in the monkey of the behaviour-based paradigm of McGonigle (1992; see also DeLillo, 1994; McGonigle and Chalmers, 1993, 1996) has supported their suggestion that a cognitive agent would demonstrate data reducing strategies such as classification and chunking in order to achieve cognitive economy in a complex serial -ordering task. Entirely success-based, individual monkey subjects have, over a period of some two and a half years, now moved through a series of increasingly demanding tasks, starting from simple discriminations,

through rule-based learning, and gradually moving towards serial-order tasks involving unprecedentedly large search spaces. The results from this series of experiments so far reported, have included the establishment of at two discrete three-item 'core' sequences and their subsequent expansions to include a nine-item string for serial order control for each monkey so far tested in the laboratory under supervised conditions. Using an incremental procedure, each subject self-qualified by success for subsequent phases which involved their solving tasks of an increasingly difficult and complex nature. The initial three-item sequence for each subject was trained with explicit order requirements and consolidated to a high level of accuracy. Subsequently, for three different four-item arrays using two exemplars with each from any two of their three categories (Phase 5), the monkeys showed that their competence for serial ordering could be maintained independent of compositionality of the sequence as a whole. Moving later to an interspersed four-condition, six-item phase, sequences again of different compositionalities were successfully reported for which the potential sequence position for a given icon had become entirely independent of its category ordinal position. Further, given the increasing probability of making an error with increasing the set-size to be searched, it was noticeable that the error-types reported at this stage were beginning to suggest much conservation of class ordering to be taking place. Indeed, it became increasingly likely after this time that more and more of the overall errors in subsequent phases would be mainly accounted for by only one or two of the one-step forwards-errors comitted at the category boundaries. All things being equal, and certainly by chance alone, one should otherwise expect error-types to become more equally represented and greater in number as the set-size of the array increased. Phase 6 saw the very first demonstrations by a non-human subject of robust, highly controlled serial order production of six-item strings, spontaneously classifying principally ordered sets of four interspersed compositionalities. Again, for each individual icon in a given six-item condition, serial position and ordinal position might vary and so the subject could not rely upon string-length alone to offer the cue to successful serial production for the array. No explanation from associative-learning theory could now be elaborated enough so as to account for this or further discoveries. Further, it was at this point also that there emerged (again within the same subject) differential reaction-time profiles for successfully produced sequences of the same length. As

could be clearly seen from the interspersed six- and ten-item conditions (Phases 6 and 9 respectively) this RT variation reliably showed phrasing effects in accordance with the category boundaries. Indeed, Bonferoni post-hoc evaluative pairwise comparisons for adjacent sequence touches showed that significantly increasing RTs most often occurred only at the category boundaries - as differentially executed by the same subject inthe same session with strings of equivalent length varying only in their categorical compositionality. These incidences of spontaneous phrasing of the items in the array (unrelated to the requirements of the task at hand) continued to be evident to some extent right up to the reaching of the twelve-item set, and each monkey continued to impose their classificatory scheme in assisting their serial order productions of a series of increasingly complex, multi-exemplar three-category extensions. It has been seen earlier in our discussion that the number of possible pathways increases exponentially as the set-size to be searched increases (see also fig. 3.2), and so the degree to which the monkey has successfully managed and controlled sequences of increasing string length (in the face of greatly increasing task difficulty) may be confirmed from the relative decline in the number of trials and errors required to reach criterion for the later phases. Figure 4.25 shows these measures for the same monkey (this time, Charlie) who not only shows a lack of dramatically increasing effort being required to cope with increasing string-lengths, but indeed took fewer trials to successfully reach an even greater level of competence with twelve items [AAAABBBBCCCC] than did he with his initial three-item [ABC] sequence acquisition.

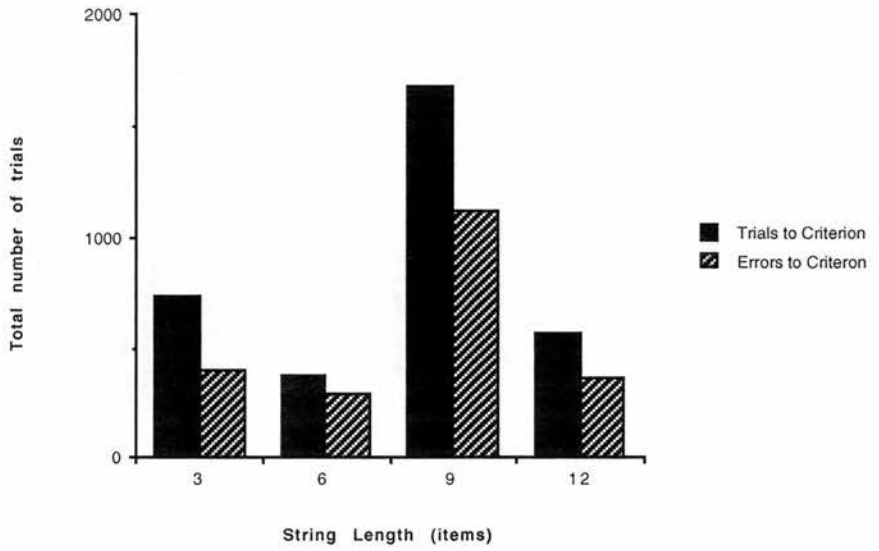


Figure 4.25 Numbers of trials to reach criterion for increasing string length in a Cebus monkey (Charlie)

Using these very novel experimental designs and procedures, it has so far been revealed that all of the *Cebus apella* monkeys so far tested have been quite capable of reliably reporting categorical seriation of 12-item sequences comprising three categories, each with four identical exemplars, at least under carefully supervised learning conditions. No monkey subject has to date failed to achieve successful seriation with any single phase condition. Another measure of the effects of the animal having informed itself (as well as having informed us) about the nature of the serial search task, and not merely having learnt to respond differentially to the absolute properties of the stimuli array, was to have seen how the monkey might transfer "what" it had learnt to a new situation involving equal numbers of previously unseen icons. Again, taking the criterion performance of the same subject (this time Alfie) the numbers of both trials and errors for the two 'core' category sequences initial-[ABC] and transfer-[DEF] for three-, six- and nine-item sets, showed clear generalisation and good class inclusion abilities to be demonstrable with novel arrays relatively quickly (see figure 4.26). RT profiles for monkeys were again different for all three sequence lengths, but this changed idiosyncratically for each subject. Whilst each continued to reliably significant category boundary effects (compared with other less salient sequence positions), some of the multiplexed [D..E] were in the opposite direction to those seen in the [A..B] conditions for the first

boundary although this was not true of the second. Some suggestion is offered that the monkey was later demonstrating the ability to search according to self-imposed within-category orders (clearly modulating the RT profiles in the case of the equivalence phases) and even larger four-item 'chunks', transcending the category boundary having gained expertise at longer sequences despite predictable classification options.

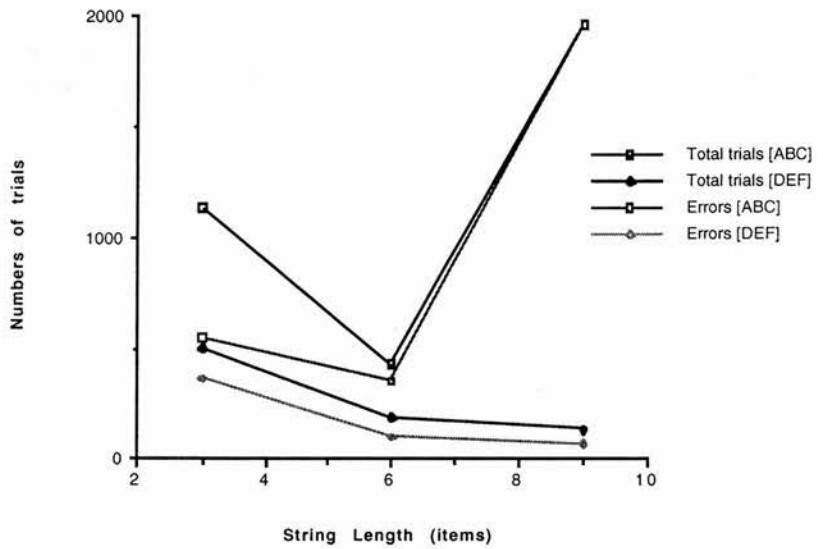


Figure 4.26 Numbers of trials and errors required to reach criterion for increasing string length: for a first [ABC] and second (transfer) [DEF] 'core' sequence in a Cebus monkey (Alfie)

Although there was no indication that twelve-item arrays may be close to an upper limit of the possible string lengths manageable by the monkey under these conditions, evidence has also been acquired to support the idea that, using such non-linguistic tasks, in an operating environment optimised to investigate the long-term growth and development of animal cognition, the first indications have also been seen of auto-regulatory behaviour in a serial, exhaustive search task. Following the expression of class inclusion management and categorical boundary effects, the monkeys have also self-imposed serial touch pattern constraints upon themselves in controlled and principled ways, well beyond the requirements of the experimental task as given. In the equivalence phases, both Charlie and Mimi showed some tendency to touch their category's size exemplars in a spontaneously ordered manner. Although not consistent across categories, each monkey repeatedly demonstrated the mergence of monotonic ordering strategies, at least in part. For the latter size-variation nine-item phase, completion occurred following as

little as one third the time taken for the colour-equivalence phase, itself already having shown a much reduced effort compared to their solving of the initial [AAABBBCCC] condition (see figure 4.27).

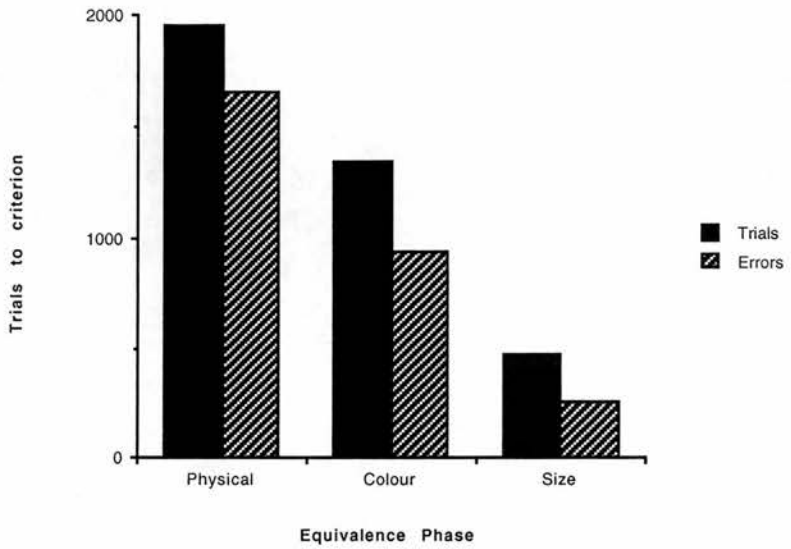


Figure 4.27 Numbers of trials and errors required to reach criterion for two nine-item sequences: for (a) colour- and (b) size-equivalence in a Cebus monkey (Mimi)

Much of the error recorded was due to single-step forwards-error which would always cause a trial to end with the subject awaiting the next array due to (intentional) programming requirements. It was always possible that the monkey might have been able to 'repair' exhaustive search by returning to the 'missed' icons, perhaps left as a result of some salience being provided by the specific configuration in view, one will never know. However, one further indication of self-regulation on the part of the subject was later to be revealed during the extended six-item, six-category 'accidental' free search task with combined [ABC] and [DEF] phases. Expecting the monkey to perhaps at best touch [ABC] then [DEF], or the converse, what both subjects went on to actually produce, were collapsed duplets composed of order-invariant icons from each of the 'core' sequences. Able to interrogate the array in any way that they wished (excepting for a reiterative touch), both monkeys proceeded to immediately, and consistently, touch the icons in orders [DA EB CF] for Charlie, and [DA BE CF] in the case of Mimi with very little deviation and few errors. Further, each produced an RT profile consistent not with duplet chunking (as seen with their previous [AABBCC] phase) but more

consistent with a chunking of an initial four-item plus a duplet to end. This last rather serendipitous finding on our 'voyage of discovery' so far, has provided very strong evidence for the existence of self-regulatory strategies in the experienced *Cebus apella* monkey interpreted being motivated by their need to cope with increasingly longer, more complex sequences of novel compositionality, using such on-line data reducing strategies in an attempt to successfully reduce an otherwise too greater demand upon working-memorial processes. Such findings continue to confirm that there is indeed a novel method which has been developed and successfully implemented here, which makes possible the capturing of a dynamic-interactive epistemological growth in the monkey subject, following its years of preparation and patient step-by-step evaluation of competences thought to underlay cognitive growth and development.

It is now time to return to our voyage, and to pursue further the degree to which the *Cebus apella* monkey might continue to inform us concerning its ability to form and demonstrate a more hierarchical classification.

Chapter 5

Classification and Hierarchical Organisation in *Cebus apella:* The study

Part 2: Two-level hierarchical grouping principles in the control of serial order production.

Introduction

Now in their third year of the experimental series, all six of the monkeys entering this series of experiments are still participating and each has been exposed to at least one [ABC] three-category, three exemplar, nine-item set. All monkeys have further provided successful demonstrations of physical equivalence, colour, and size variation phases for all categories, and a second [DEF] transfer-set has also been seen by half of them. All of the monkey subjects have shown the emergence of cognitive strategies as a result of the increasing task demands, as indexed by successful classification (chunking) in a task which required the categorical seriation of multi-exemplar sequences varying in compositionality. This was all the more impressive with increasingly difficult search problems as the set-size increased in the face of the combinatorically explosive space of possible paths made available through the use of larger sets. Having recorded the successful exploitation of a simple one-level classificatory scheme and consistent positive transfer effects both within classes and for extended sequences comprising increased numbers of icons, it was fast becoming apparent that these novel sequences were being 'solved' relatively easily, despite the combinatorically explosive circumstances involved. However, simple classification of this sort would not continue to be so effective when tackling very much longer sequences for which a category membership itself became too big, resulting in an effectively large search problem of its own. One solution to such a problem would be to achieve some kind of control via a hierarchically organised principle which might conceivably keep the search within more manageable proportions. Having fulfilled the prerequisite, qualifying levels of preformance with a nine-itam series, and in keeping with the proposed extendibility options for either breadth or depth which an otherwise linearly increasing sequence might provide (see fig. 3.3), by now introducing levels of physical variation amongst the

exemplars of each class once more, one could assess their achievements further by imposing *within*-class ordering requirements. Thus, within the same experiment, classical seriation was to be combined with classification, in a task for which the explicit seriation of individual category exemplars and simultaneous seriation of all classes would, if successful, give rise to the production of the first recorded account of a two-level hierarchical control mechanism in the non-human primate.

In taking this new step in our voyage of discovery, we will be travelling with the knowledge gained from our subjects previous history of successes (and failures) of their control and production of serial order search tasks. These novel fixed-order tasks would again be increasing the levels of difficulty as the string-length increases, but this time, with each incremental step, the number of possible error-paths increases at an even higher rate now that a single, unique solution will be required of the subject for any given test condition. With the stakes now raised so high, this series of experiments will help us determine whether the monkey will continue to successfully control and report unique serial order productions of progressively longer sequences for as long as classification remains an option. Further, we shall be able to determine the degree of spontaneity with which monkey subjects search for items that are class inclusive, rather than opt for those in different classes, whilst finding such successes to be associated with accelerating acquisition rates.

Simplified design and method.

Icon sequencing characteristics and reaction time data were recorded for subjects working through acquisition, classification and seriation of up to nine-item arrays with explicit ordering requirements, presented on a computer touch-screen apparatus as detailed above. A total of 22 consecutive phases comprised the principle two-level hierarchical phases of the experimental series of the study to date. Individuals self-qualified by success for inclusion in subsequent phases of the experiment using a longitudinal within subject design.

Subjects

Six sexually mature adult Cebus apella monkeys have so far participated in this extension of the study, two male, four female. Two of them (Kissy and Ollie) had only recently joined this stage, and did not therefore contribute data to this part of the report. A considerable laboratory housing refurbishment had taken place prior to any monkey starting this phase of the experimental series, but all of them were either working at, or near, nine-item sequences immediately prior to starting the two-level hierarchical phases (for 'refresher' reasons or because they were still to complete their previous equivalence phase trials following the laboratory's experimental restart period). All of the monkeys were by now an estimated 13-14 years old, and had at least three years of experience with the apparatus and serial search tasks, all successfully having completed at least four different nine-item sequences in the laboratory under supervised conditions.

Stimuli

As stimulus material, two nine-item sequences comprising three familiar shape categories (the monkey's own original [ABC] or [DEF] sets), each with known discriminable and explicitly orderable exemplars (three colours and three sizes of each icon shape) could be randomly positioned and presented within a symmetrical five by four matrix (170 x 190 mm) on the surface of the interactive video touch screen as before. Each subject had already been assigned a unique colour sequence (e.g. red square [A], green circle [B] & blue triangle [C]) and this same colour sequence was to be used *within* each class, so conserving, whilst at the same time removing another potential variable in controlling for stimulus-preference effects across animals. The initial three sizes of each icon class were identical to those previously used for the earlier [ABC] size-equivalence phases, being seen here as white outline contour shapes on a black background.

Apparatus

The experimental test environment used for all phases of this study was as previously described, but with updated in-house purpose-written software, and IBM-compatible 486 PCs replacing the previous computer equipment

employed to both generate the experiments and to automatically record the touch-screen response data in real time.

General Procedure.

Having voluntarily withdrawn directly from its newly refurbished family-grouped colony room and situated in wheeled transit housing, each monkey was tested daily as before at their familiar fixed location. The monkeys continued to work in the same pairs inside the purpose-built testing suite in the laboratory, situated at adjacent testing areas within visual and auditory range of each other. Lighting was subdued so as to provide better focus towards the touch-screen environment and a white-noise generator was operating to filter out extraneous auditory inputs from the nearby monitoring suite and offices. All subjects, once comfortable in the test environment, were presented with a number of icons upon the touch-screen monitor and were required to make serial exhaustive touch responses towards them with an outreached arm. The novel aspect of this particular series of experiments was that serial exhaustive search should continue to maintain category order, but also to order every icon *within* any given category. This was to be completed in an explicitly ordered manner, irrespective of the number of icons presented (i.e., a unique x-item order would always be required for any given condition). At no time throughout the experimental series was the subject given any explicit tuition, differential reinforcement, or had other experimenter interaction (although sometimes present). Having touched a target icon, a tone would sound and the target would be highlighted, then disappear from view for 200 mS. For all screen displays in the experimental series, after each of the targets had been correctly touched, the array was removed from the screen which blanked for 14 seconds. If, and only if, the touch sequence showed neither reiteration nor forwards-error touch of any of the target items to be ordered in the array, a single peanut was delivered by the dispenser. Repeated touching of a given icon prior to moving on to the next was permitted. If, however, a subject showed a reiteration or forwards-error of target touch responses for the current screen array, the screen became blanked white for 30 seconds, the feed-dispenser remaining inactive. The inter-trial interval for correct trials was set at 14 seconds, the whole session capable of automatic completion either following a given number of trials or until

the reaching of the experimental criterion pre-set for the phase. Whether or not the monkey had successfully completed any given trial, each subsequent trial presented the icons of the same phase's conditions in a new configuration, each to one of the twenty possible positions determined by a pseudo-random number generated automatically by the computer program. Subjects could be taken through sequences of incremental item numbers and combinations of category arrays in a series, up to and including a maximum of 20 items per screen array, dependant upon their degree of success and task-motivation as determined by the monkey's successive phase criteria performances. The entire 22-phase experimental series comprised four principal sub-divisions:

(a) 3-class, 3-sizes - categorical seriation - [acquisition phases 24-31]
(b) 3-class, 3-colours - categorical seriation - [acquisition phases 32-39]
(c) categorical versus linear seriation (6-sizes) - [phases 40-43]
(d) categorical versus linear seriation (9-sizes) - [phases 44-45]

A subject would receive remedial trials as appropriate after stasis or a declining performance over 10 sessions, and this would typically involve a return to the previous phase completed, once again to satisfy the criterion level of performance before continuing with the next phase once more. At no time was differential feedback offered during the entire experimental series. Each monkey was run within the same subject-pairing in a pre-determined and fixed order, for a maximum or 40 minutes at the same hour of the morning each day, 5 days a week. The experiments were automated for both stimulus presentation and feeder access components in daily sessions, each containing a maximum of 50 trials per subject per day unless a phase criterion was met within the session.

In accordance with the earlier rationale and figure 3.3 above, the following Tables 5.1-5.23 show the individual phase results and qualifying conditions for the participation of each monkey in the consecutive experimental test conditions of phases 24 through 44 to date. These were to be completed by each monkey in accordance with the specific phase procedures as they evolved, and are listed below, each in the order in which they were presented. Although all six monkeys were to enter the experimental series from the outset, it was not possible to predict in advance how many of the monkeys might succeed at any given level of

difficulty as the sequences to be interrogated became longer and more complex. Because each animal would qualify for its inclusion in a particular phase having reached criteria level performance at the previous phase, descriptive results including the numbers of trials, errors and percentage errors to criteria will be given for each phase. By so doing, it will quickly become evident how each monkey's qualifications for continuation through the experimental series were to develop, both longitudinally and with respect to the performance of the other monkeys participating in the study. As a general code, the following key provides examples of the nomenclature to be used to identify string composition throughout the following tables:

For example: for a given sequence array $[A_{c1}A_{c2}A_{c3}B_{c1}C_{c2}]$:

A_{c1}= 1st category, 1st coloured icon in 1st ordinal position (e.g., red star)
A_{c2}= 1st category, 2nd coloured icon in 2nd ordinal position (e.g., yellow star)
A_{c3}= 1st category, 3rd coloured icon in 3rd ordinal position (e.g., blue star)
B_{c1}= 2nd category, 1st coloured icon in 4th ordinal position (e.g., red hexagon)
C_{c2}= 3rd category, 2nd coloured icon in 5th ordinal position (e.g., yellow square)

and similarly for size stimuli, e.g. $[A_{s1}B_{s2}C_{s3}]$:

A_{s1}= 1st category, 1st sized icon in 1st position (e.g., small, white outline star)
B_{s2}= 2nd category, 2nd sized icon in 2nd position (medium, white outline square)
C_{s3}= 3rd category, 2nd sized icon in 3rd position (large, white outline circle)

Specific phase details and procedures.

(a) 2-level categorical (size) seriation - acquisition phases

Introduction

The conditions of the first two-level hierarchical seriation experiments, Phases 24 through 31 are shown in Table 5.1 below and include the details of interspersed control conditions and completion criteria for each phase. All of the monkeys participating in this experimental series would have started with Phase 24, but at the time of the first monkey's attempts there was yet to be any well developed and detailed procedures for these phases. As a result, five of the subjects started this experimental series with what is now Phase 25, and one other started with Phase 26. Thereafter, most of

the monkeys were to follow the subsequent phase steps in the order as listed, once the sequence increments had been established. Charlie was the only subject to have completed three, six, and nine-item sequences in this series without the intervening incremental stages prior to the new design being agreed upon. All phases required an extremely high level of vigilance and unusually consistent performance in the face of increasing task difficulty. Task difficulty here was not only increasing with respect to the combinatorial explosion of possible pathways through increasing search spaces with larger string-lengths (see fig. 3.2), but were further to be exacerbated by there now being a far greater numbers of error-types for any given string-length than before, now that the explicit ordering of each and every icon of the set was required. Another feature of this incremental series was that the novel icon to be added when each time increasing string-length would be inserted at different points of the sequence, depending upon the nature of the last string composition most recently completed. Thus a new string-length would not simply result from the addition of an extra icon to the end of the previous string. For all subsequent phases, individual monkeys were self-selecting by their continuing successful attainment of the previously given phase criteria as they progressed through tasks of increasing string length and thus greater levels of difficulty.

Table 5.1 Size order Phases 24-31 [ABC]
2-level hierarchy: 3 sizes/3 categories

Phase No.	Fixed order Conditions	Completion criterion
24	$A_{s1}A_{s2}A_{s3}$ $B_{s1}B_{s2}B_{s3}$ $C_{s1}C_{s2}C_{s3}$	15/20 15/20 15/20
25	$A_{s2}B_{s2}C_{s2}$	15/20
26	$A_{s2}B_{s2}C_{s2}$ $A_{s1}A_{s2}B_{s2}C_{s2}$	3/4 15/20
27	$A_{s1}A_{s2}B_{s2}C_{s2}$ $A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}$	3/4 15/20
28	$A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}$ $A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}$	3/4 15/20
29	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}$ $A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s2}$	3/4 15/20
30	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s2}$ $A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}$	3/4 15/20
31	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}$ $A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}C_{s3}$	3/4 15/20

The actual sequence compositions presented during these acquisition phases were drawn from a larger number of possible decomposition strings derivable from an [ABC] icon set, and were those conceivably simpler in structure by comparison. For every phase condition in this later series a unique category and category exemplar order was required to be reported by each subject for each icon set in the array of every trial. Thus, for successful completion of all future phases, every new task demanded an explicit and unique ordering of each and every individually identifiable icon upon the screen array. From trial to trial the same icons were presented in different configurations so as to avoid the production of stereotyped motor responses, whilst at the same time provoking continuous array interrogation on the part of the subject. The use of an interspersed control condition (the test condition of the previous string length and compositionality) was also given partly for this reason, so preventing the subject's responses becoming too reflexive and thereby less 'cognitive' in production, whilst simultaneously allowing interpretation of any significant amounts of error in the face of the greatly increasing task demands under supervised learning conditions.

Phase 24

Rationale. To allow the demonstration of the existence of a base-line performance from which any depth of search might be deemed supportable, this first of the 2-level (size) hierarchy phases was given so as to ensure that the pre-existing requirements be satisfied for the subject being able to order the individual icons of each category. This first phase would further inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli in a consistent and orderly manner, independantly and simultaneously for each class.

NB: This phase would only be given to those subjects who had not recently been exposed to these same stimuli in their size-equivalence phases. However, the procedure is given in full below for completeness.

Subjects. No subjects have to date had need to take part in this phase.

Design and Procedure. In accordance with the general procedure above, this phase presented three conditions $[A_{s1}A_{s2}A_{s3}]$, $[B_{s1}B_{s2}B_{s3}]$ and

[$C_{S1}C_{S2}C_{S3}$], each of which required that their three icons (all monochrome contours) of the categories [A], [B] or [C] be reported in order of increasing size. Each category's exemplars were to be interrogated and touched sequentially - [X_{S1}] first, followed by [X_{S2}], followed by [X_{S3}] (where X=category marker) - thus requiring of the subject a unique solution to be found for every array presented. Four error-types were possible (three forwards errors and one reiteration), repeated touches to the last touched icon still being permissible. The individual conditions were given in a random order over 50 trials on a ratio schedule of 1:1:1. Phase criteria required the subject to complete fifteen correct trials out of the last (cumulative) twenty for each condition. Upon reaching a condition criterion, that particular condition would not be presented again during the phase.

Phase 25

Rationale. To permit the firm establishment of the new 'core' sequence for subsequent generalisation and extension, a single test condition [$A_{S2}B_{S2}C_{S2}$] using three previously seen stimuli, (the original [ABC]-monochrome contour set) were presented for discrimination and ordering. This first phase would again inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli in a robust and fixed-order.

Subjects. Three adult monkeys (Charlie, Alfie and Luba) took part in this phase.

Design and Procedure. In accordance with the general procedure as before, this phase presented a single condition [$A_{S2}B_{S2}C_{S2}$] with all subjects using three icons of equal size, but discriminable by shape and spatial location shown simultaneously to the touch-screen. These icons were to be interrogated and touched sequentially - [A_{S2}] first, followed by [B_{S2}], then finally [C_{S2}], requiring of the subject a unique solution to be found. Four error-types were possible (three forwards errors and one reiteration), with repeated touches to the last touched icon still being permissible. The subject completed this phase by achieving fifteen correct out of the last (cumulative) twenty trials. To assure the establishment of this 'core' [$A_{S2}B_{S2}C_{S2}$] sequence, the phase was consolidated by each

subject being required to reach the task criterion twice before moving on to the next string in the incremental series.

Results. All three monkeys completed this condition and qualified for continuation to the next phase of the two-level hierarchical series of arrays. The mean number of trials, errors and percentage errors to criterion for the condition is shown in Table 5.2.

Table 5.2 Nos. of trials, errors and % errors to criterion [$A_{s2}B_{s2}C_{s2}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	18	3	16.7%
Alfie	63	34	53.9%
Luba	15	0	0.0%

All subjects completed this new 'core' sequence acquisition condition almost immediately despite the higher level of performance required, and thus showed themselves capable of reporting yet another novel three-item, (three category) sequence. Only two subjects showed any error, and all were of one type (a touch to [B_{S2}] first). Reaction time (RT) profiles for the last fifteen correct trials (those for which the most consistent and reliable data may be obtained) indicated significantly decreasing times to be taken by all of the monkeys to touch each subsequent icon once they had begun to interrogate the touch-screen array ($p<0.05$). Between-subject variation was very low, and pooled RT mean data for the six animals was as shown in Figure 5.1.

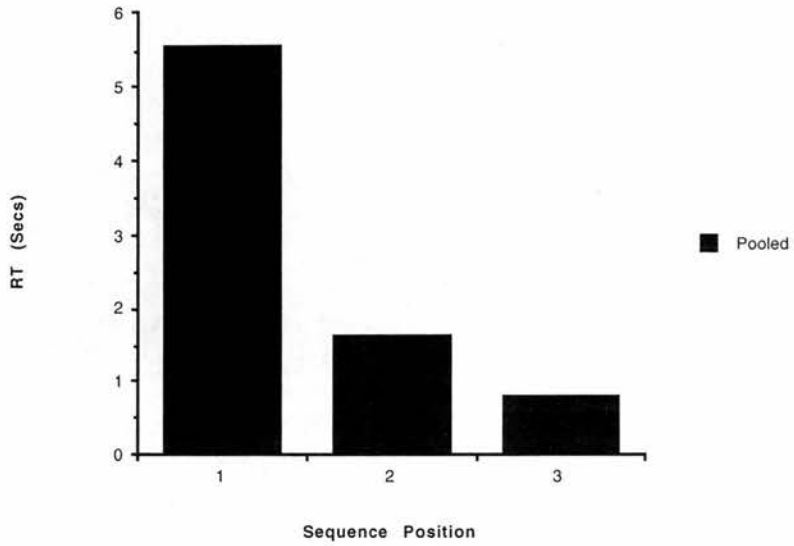


Figure 5.1 Mean RT profile for successful 3-item, three category fixed-order (size) sequencing by Cebus monkeys (N=3, pooled)

Discussion. It had now been established that each monkey would consistently report their new 'core' three-item sequences following a supervised training procedure which allowed no flexibility of ordering on the part of the subject. Compared to the earlier [ABC] and [DEF] training phases, it can be seen that this new three item sequence was 'seen' almost immediately as a permissible variation of the previous [ABC] set. For the pooled data, all of the errors recorded were of one type, [B_S2] with close inspection of video-cassette tape usually revealing poor screen-touching (and hence bad registration) by the subjects. There was no evidence of sequential search ambiguity or difficulty. The immediate high levels of success (Luba=100%, Charlie=83%) and the absence of three out of four possible error typologies was suggestive of all monkeys demonstrating high degrees of conservation of class ordering from their earlier training phases (although familiar icons, they had never been grouped together in this way before in such a task). This idea may purchase some support from the RT findings of all six monkeys, in which the time taken to search each subsequent icon of the set to be ordered, significantly decreased throughout sequence execution, suggestive of the development of a route-planning strategy to be taking place. The initial reaction times (IRTs) were unusually large here compared to those for recent performances with strings of much greater length.

Phase 26

Rationale. With a view to determining whether the monkeys were going to be capable of continued spontaneous classification of known familiar icons in an array, the next phase set out to explore the possibility of serial order information transfer to a novel four-item set, by increasing the sequence length by one icon, this time adding the need for explicit *within*-category orderings as well as maintenance of the existing class order. This novel core-sequence expansion would not only offer an indication of the monkey's ability to spontaneously classify a principally orderable set, but would also, if successful, provide us with the first examples of a non-human primate to be capable of working with a hierarchical grouping principle, albeit under supervised learning conditions. Indications of the monkey's having derived any early success as a result of the use of data reducing strategies would be clear from accelerating acquisition rates given the simultaneous increases in sequence length and level of difficulty that this phase provides.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the first of an incremental set of phases for which the previous test condition was to be raised in string length by one icon in order to extend the sequence to be searched in such a way as to produce a nine-item string with three-categories, each to contain three orderable exemplars within them. This first increment saw the addition of an icon to the first ordinal position and thereby presents two interspersed conditions of $[A_{S}2B_{S}2C_{S}2]$ and $[A_{S}1A_{S}2B_{S}2C_{S}2]$. The two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the first category's smaller sized icon to the test condition of the previous phase. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All of the monkeys completed this phase, achieving the 75% criterion level required, whilst simultaneously maintaining a high standard of performance with the control condition. The mean number of

trials, errors and percentage errors to criteria are shown in Table 5.3.

Table 5.3 Nos. of trials, errors and % errors to criterion [$A_{s1}A_{s2}B_{s2}C_{s2}$]

Subject	No. of Trials	No. of Errors	% Errors
Mimi	143	96	67.1%
Alfie	582	363	62.4%
Luba	341	212	22.2%

All of the subjects showed immediate successes with between 25%-35% correct trials out of their first twenty, and analysis of the incorrect trials revealed quite a consistent pattern to be found, in which unequal representation of the nine possible error types was evident. Accounting for between 89% and 100% of the total error trials completed for any one subject in this phase, it was found that despite there being nine possible error paths executable, most could be accounted for by only two of them. Whereas Alfie's error profile showed dominantly [A_{S2}]-1st touch errors, for the other two monkeys the dominant error type reported was [$A_{S1}B_{S2}$] (55%), a one-step forwards-error, followed by [A_{S2}]- 1st touch error (18.4%). RT analysis (figure 5.2) revealed decreasing times for each subsequent touch, but no significant category boundary effects were in evidence for any subject. Equal variance allowing pooling of data to show a similar profile to that seen in the previous phase.

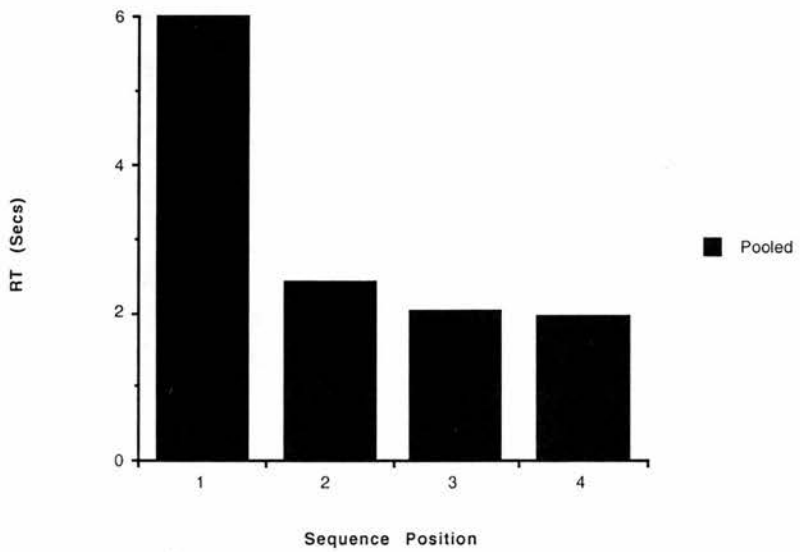


Figure 5.2 Mean RT profile for successful 4-item, three category fixed-order (size) sequencing by Cebus monkeys (N=3, pooled)

Discussion. The criteria performance measures have shown that all of the monkeys would successfully report a further four-item with novel compositionality, using all three category markers with each maintaining their relative sequence order with two fixed-order exemplars in the first two ordinal positions. Early successes were shown by all subjects (each with between two and six correct in the first ten trials) and despite the increased difficulty, each completed the phase after taking only a few more trials and errors than were previously required for them to complete their last three-item sets. First touch errors were recorded in unusually high numbers, especially salient since this might reflect some initial neglect of the new icon, which was to be inserted into the first position of the existing sequence. Despite the possibility of some nine error types for the test condition now, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 80% of the remaining error trials for which this strategy did not allow serial success. This latter error-type was the now familiar one-step forwards-error, the subject moving on to start the next category before having touched all exemplars of the last. Although no boundary effects were revealed by analysis of any subject's RT data, increased IRTs were in evidence once more, possibly suggesting greater planning and memory being required for the four subsequent explicitly ordered touches to be made.

Phase 27

Rationale. Having by now satisfactorily demonstrated the ability to negotiate a four-item fixed-order sequence composed of three categories, the first with two exemplars, it was now reasonable to give the monkey another increase in string length following the addition of a further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order information to a novel five-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and level of difficulty, indications of the monkey's having derived any early success as a result of the use of data reducing strategies would be clear from accelerating acquisition rates and the development of any category boundary effects for individual subject

RT profiles. It was to be of particular interest to note whether category boundary effects would re-emerge in the presence of this three-item within category order requirement condition, now that the sequence length was approaching the 4-5 item limit of old.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the second of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the third ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{s1}A_{s2}B_{s2}C_{s2}]$ and $[A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}]$. The two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the first category's larger sized icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this phase, achieving the 75% criterion level required for the test condition, whilst simultaneously maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.4.

Table 5.4 Nos. of trials, errors and % errors to criterion $[A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}]$

Subject	No. of Trials	No. of Errors	% Errors
Mimi	169	95	56.2%
Alfie	333	239	71.8%
Luba	473	283	59.8%

All subjects showed immediate successes with two monkeys showing at least 35% correct trials out of their first twenty. Analysis of incorrect trials revealed quite a consistent pattern to be found across subjects, in which unequal representation of the eighteen possible error typologies was evident. For as much as 87% of the total error trials completed for any one subject in this phase, it was found that the same two error paths could

account for the incorrect touches made. Whereas Alfie's error profile continued to show dominantly [A_{S1}]-1st touch errors, for the other two monkeys the dominant error typology reported was [$A_{S1}A_{S2}B_{S2}$] (40%), a one-step category boundary forwards-error, followed by [$A_{S1}A_{S3}$] a one-step within-category forwards-error (35%). RT analysis revealed only one significant category boundary effect, which for case of Mimi, showed a significantly marked rise in the search time only for the last category icon ($p< 0.05$) as may be seen from figure 5.3. For Alfie, noticeable, but non-significant rises were seen for both the third and final sequence positions (see figure 5.4). The third subject (Luba) showed non-significant steadily declining reaction times with each subsequent touch in a manner similar to that seen in the previous phase.

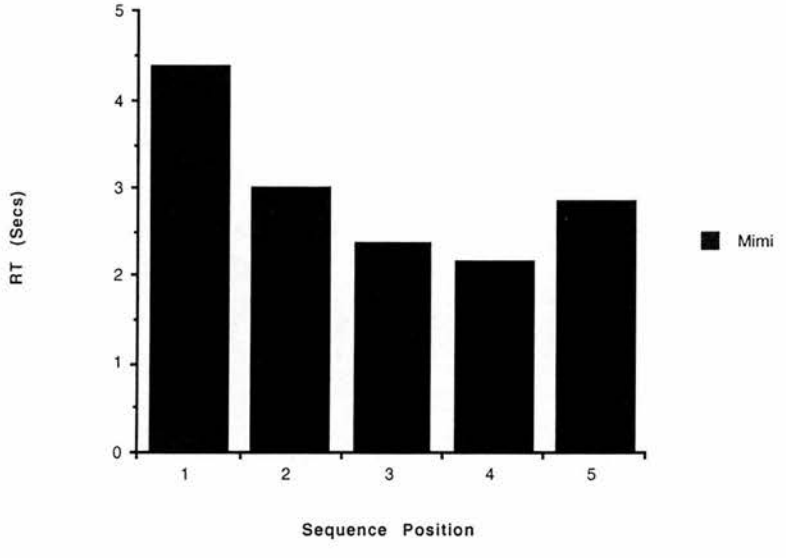


Figure 5.3 Mean RT profile for successful 5-item, three category [$A_{S1}A_{S2}A_{S3}B_{S2}C_{S2}$] sequencing by a Cebus monkey (Mimi)

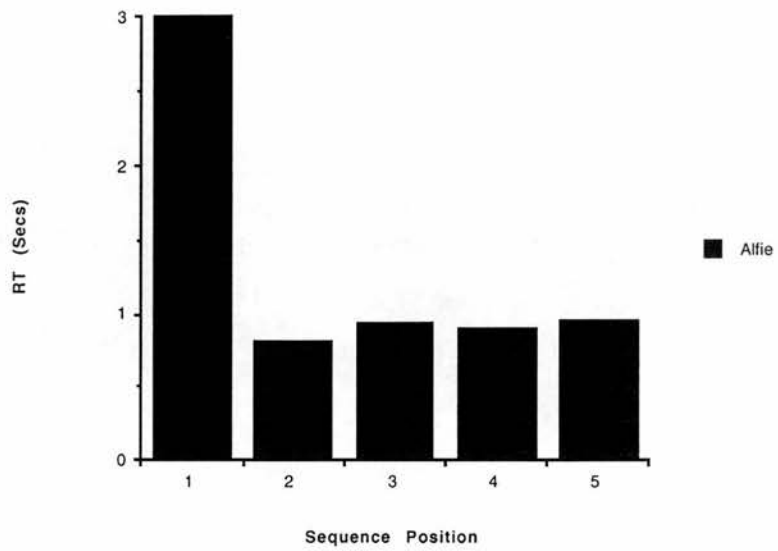


Figure 5.4 Mean RT profile for successful 5-item, three category fixed-order (size) sequencing by a Cebus monkey (Alfie)

Discussion. This phase has demonstrated for the first time that the monkeys would successfully report a five-item series, using three category markers maintained in a relative sequence order, together with three exemplars to be reported in a fixed order for the first three ordinal positions. Early successes were shown by all subjects (each with between three and six correct of the first ten trials) and despite the increased difficulty, each completed the phase after taking fewer trials and errors than were previously required for them to complete their last three-item sets. First touch errors were recorded in unusually high numbers for one subject as before, mainly as a result of poor screen touching. Despite the possibility of some eighteen error types for the test condition now, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 50% of the remaining error trials for which this strategy did not allow serial success. The latter error-types were both the familiar one-step forwards-error, the subject moving on to start the next category before having touched all exemplars of the last. Although boundary effects were revealed by analysis of two subject's RT data, only one was significant, for the final category. The phrasing otherwise varied idiosyncratically for each subject with Alfie showing a tendency to 'chunk' the array into a two-plus-three sequence execution, $[A_{S1}A_{S2}]$ followed by $[A_{S3}B_{S2}C_{S2}]$ ignoring the more usual categorical phrasing pattern. A second subject,

Mimi, showed a preferred four-plus-one arrangement, and the remaining monkey showed a systematic, but non-significant reduction in the times taken to touch the icons as the sequence progressed. We see here for the first time that although all monkeys are demonstrating clear generalisation and transfer of their previously learned sequencing skills and good evidence for the existence and use of some strategic planning (or at least economic management) with a novel larger sequence, each monkey can do so in a way which does not express itself in a stereotyped way when classification options are presented under supervised conditions. It would be interesting to see in what ways this particular sequence might have been exhaustively searched under free-search conditions, but for now our first question requires that fixed-ordering be continued for increasingly larger sets.

Phase 28

Rationale. Having by now satisfactorily demonstrated the ability to negotiate a five-item fixed-order sequence composed of three categories, the first with three exemplars, it was now possible to give the monkey a further increase in string length following the addition of a further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order information to a novel six-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and level of difficulty, it was also to be of interest to see in what ways the category boundary effects might be modulated in the presence of one three-item, and a second two-item, within-category order requirement condition.

Subjects. Three adult monkeys (Alfie, Mimi and Luba) took part in this phase, although the data from a fourth subject (Charlie) is included here despite his six-item phase compositionality being different.

Design and Procedure. This phase saw the third of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the fourth ordinal position with the previous test condition now serving as a

control, the phases thereby providing two interspersed conditions of $[A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}]$ and $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}]$. In the case of a fourth subject, Charlie, a six-item condition comprised icons from only the first and second categories giving a $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}]$ test sequence (with $[A_{s1}A_{s2}A_{s3}]$ as control condition), his size-ordering experiments having taken place prior to the establishment of the sequence of incremental phases now scheduled as above in Table 5.1. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the second category's smallest sized icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All four monkeys completed their six-item fixed-ordering phases, achieving the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.5.

Table 5.5 Nos. of trials, errors and % errors to criterion $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie *	500	327	65.4%
Mimi	214	112	52.3%
Luba	393	218	55.4%
Alfie	1312	852	64.9%

* = actual sequence composition => $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}]$

All of the subjects showed immediate successes with all except one subject showing some correct trials in their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found across subjects, for which very unequal representation of the possible error types was again evident. Accounting for between 63% and 74% of the total error trials completed for any one subject in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only four of them. Whereas Alfie's and Charlie's error profile continued to show many $[A_{s2}]$-1st touch errors, for all monkeys the dominant error typologies reported were $[A_{s1}A_{s2}B_{s2}]$ (30%) and

[$A_{S1}A_{S2}A_{S3}B_{S2}C_{S2}$] (7%), both one-step forwards-category errors, followed by [$A_{S1}A_{S3}$] (20%) and [$A_{S1}A_{S2}A_{S3}B_{S2}$] (20%), both one-step within-category touch errors. For Charlie, a noticeable further 10% of incorrect trials could be accounted for by [$A_{S1}B_{S1}$] errors. All subjects showed negligible numbers of reiterative touches. RT analysis revealed rise times only at the category boundaries (although non-significant) for each of the three subjects, with equal variance allowing data to be pooled, as shown in figure 5.5 below. For Charlie, there was no category effect noticeable at all. His RT profile showing declining times (not significant) to be taken for every subsequent touch as he progressed through his six-item two-category, three exemplar set in its pre-determined fixed-order.

Discussion. This phase has produced clear evidence that the monkey would successfully report a six-item series, using three category markers in a fixed relative sequence order, together with two or three exemplars of the first two categories required to be reported in a fixed-order in the first five ordinal positions. Early successes were shown by all subjects (with two showing at least two correct of the first ten trials) and each completed the phase after fewer trials and errors than were previously required for them to complete phases of equivalent levels of difficulty.

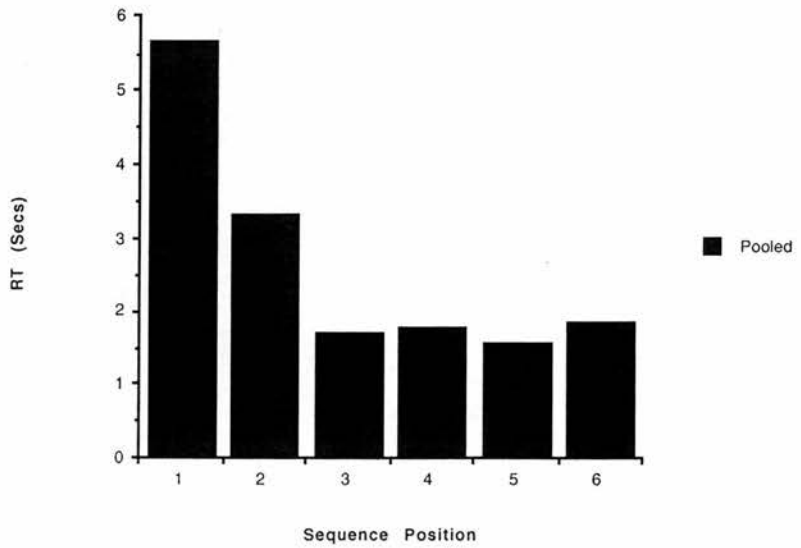


Figure 5.5 Mean RT profile for successful 6-item, three category fixed-order (size) sequencing by Cebus monkeys (N=3, pooled)

Despite the possibility of some twenty-five error types for the test condition in this phase, error profiles reveal clear preferences for

touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 60% of the remaining error trials for which this strategy did not allow serial success. These latter error-types included both the familiar one-step forwards-error, (the subject moving on to start the next category before having touched all exemplars of the last) with the addition of new type of errors resulting from the novel requirement of fixed-order sequencing. These new error-types were also one-step forwards-errors, but this time took place *within* rather than across category boundaries, and accounted for almost a third of all errors recorded. One point to note about Charlie's error profile, was the occurrence of a rather unusual quantity of errors (about 10%) being made in the second position. The particular touch error was always to the same second icon [$A_{S1}B_{S1}$] which suggests that this subject was perhaps attempting to collapse the set by his own principle (first-things-first, second-things-second, etc.., one not permitted by the programme at this time) as was seen in his combined [ABCDEF] condition earlier (see results of Phase 22 above). Although boundary effects were revealed by analysis of individual and pooled RT data, none were significant, for any category. The phrasing for all subjects, however, showed a tendency for them to 'chunk' the array into at least a three + two sequence execution, entirely consistant with the three categorical boundary positions for this sequence. A lone subject, Charlie, working with a different compositional six-item sequence, showed a systematic reduction in the time taken to touch each icon of the two serial monotonic sequences of his string. Taking such a long start up time (Mean IRT= 10.4 Secs) in comparison to his previous 9- or even 12-item sequences (typically less than 2.0 Secs), Charlie was perhaps able to 'see at a glance' what the solution could be, and planned the entire six touches to be made from the outset. We continue to see here clear evidence of generalisation and spontaneous class inclusion of novel items added to existing sequences, mid-sequence, whilst demonstrating the existence and use of strategic planning and economic management of increasingly larger fixed-order sequences of increasing complexity and difficulty.

Phase 29

Rationale. The successful completion of six-item fixed-order sequences composed of three categories, the first and second with multiple exemplars, is now further extendible, giving the monkey yet another increase in string length by the addition of a further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase assessed the degree of possible generalisation and spontaneous classification of serial information with a novel seven-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and difficulty, it was again to be of interest to see in what ways the category boundary effects might change in the presence of a simultaneous two three-item, within-category order array in the same condition.

Subjects. Three adult monkeys (Alfie, Mimi and Luba) took part in this phase. (The fourth subject, Charlie, next saw his nine-item, three category, three (size) exemplar, fixed-order set, see Phase 31 below).

Design and Procedure. This phase saw the fourth of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon into the sixth ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}]$ and $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s2}]$. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the second category's larger sized icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this novel seven-item sequence seen as a fixed-order requirement condition. All subjects achieved the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control trials. The mean number of

trials, errors and percentage errors to criteria are shown in Table 5.6.

Table 5.6 Nos. of trials, errors and % errors to criterion [$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s2}$]

Subject	No. of Trials	No. of Errors	% Errors
Mimi	151	93	61.5%
Luba	344	177	51.4%
Alfie	420	303	72.1%

All subjects showed immediate success, with each showing between 5% and 15% correct trials in their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found across subjects, for which very unequal representation of the possible error types was evident. Accounting for between 65% and 77% of the total error trials completed for any one subject in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only four of them. Alfie's [A_{S2}]-1st touch errors were now very much reduced (although not to insignificant amounts), but for all monkeys the dominant error typologies reported were [$A_{S1}A_{S2}B_{S2}$] (17%) and [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}C_{S2}$] (18%), both one-step forwards-category errors, followed by [$A_{S1}A_{S3}$] (12%) and [$A_{S1}A_{S2}A_{S3}B_{S2}$] (23%), both one-step within-category touch errors. All subjects showed only a few error trials involving any reiterative touches. RT analysis revealed no (significant) category boundary effects for any of the three subjects, although many individual trials showed strong indications of pausing at both the first category boundary, and at the sixth position prior to touching the first and only icon of the third category. The strongest (but still non-significant) effects for pausing were seen in the profile of Mimi's reaction times, although these were not coincident with either of the category boundaries, as shown in figure 5.6. For both Alfie and Luba, RT profiles showed declining times to be taken for each subsequent touch (excepting Alfie's sixth on occasion) as they progressed through their seven-item three-category, multi-exemplar set in its pre-determined fixed-order. Other post-hoc evaluative pairwise tests showed no significant rise-times for any of the individual subject's RT profiles.

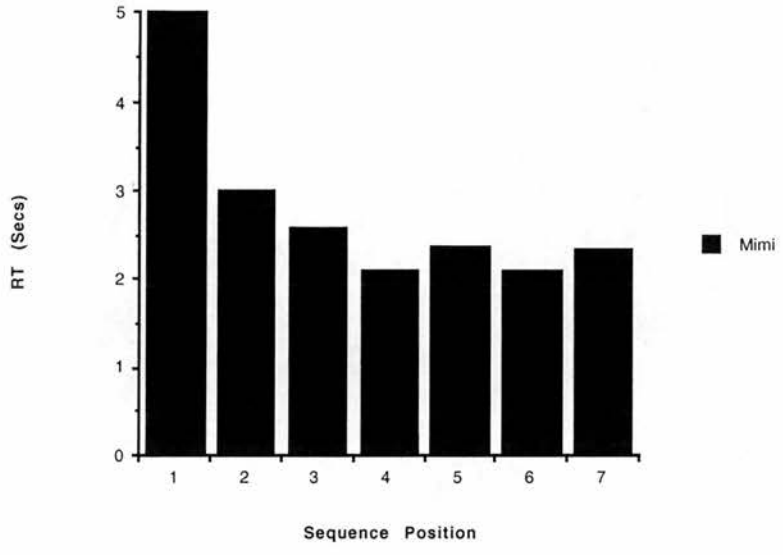


Figure 5.6 Mean RT profile for successful 7-item, three category fixed-order (size) sequencing by a Cebus monkey (Mimi)

Discussion. This phase has demonstrated for the first time that the monkey would successfully report a seven-item series, here using three category markers maintained in relative sequence order, together with the three exemplars of the first two categories to be reported in fixed-orders in the first six ordinal positions. Early successes were shown by all subjects (each with between one and six correct of the first ten trials) and despite the increased difficulty, each completed the phase after taking fewer trials and errors than were previously required for them to complete their last four-item sets. First touch errors were recorded in fewer trials than ever before, as were reiterative touches (with a total of only three for one subject). Despite the possibility of some thirty-six error types for the test condition now, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 65% of the remaining error trials for which this strategy did not allow serial success. These latter error-types were both the familiar one-step forwards-error (the subject moving on to start the next category before having touched all exemplars of the last), and the new one-step forwards-error seen within a category, now itself requiring its exemplars to be explicitly ordered. Category boundary effects were evident for selected individual trials, but were weak as revealed by analysis of pooled and individual monkey RT data for criterion runs (for which none were significant). The phrasing otherwise continued to vary

idiosyncratically for each subject with Alfie showing a tendency to 'chunk' the seven-item array into a five+two item sequence execution, and Mimi preferring a four+two+one item chunking, both ignoring the more usual categorical phrasing pattern seen earlier. The remaining monkey (Luba) continued to show a systematic reduction in the time taken to touch subsequent icons as the sequence progressed. We see here again that although all monkeys are demonstrating clear generalisation and transfer of their previously learned sequencing skills and good evidence for the existence and use of some strategic planning with increasingly larger sequences, each monkey appears to do so in a way which does not express itself in any stereotyped fashion when classification options be presented under supervised conditions. It therefore continues to be of interest to see in what ways this might carry on for increasingly larger search spaces.

Phase 30

Rationale. Having now satisfactorily demonstrated the ability to negotiate a seven-item fixed-order sequence composed of three categories, the first two with three differently sized exemplars, it was now possible to give the monkey a further increase in string length following the addition of one new icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order control to a novel eight-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing categorical order. Increasing the search set both in complexity and difficulty, it was to be of interest to see in what ways the category boundary effect might continue to alter in the presence of two three-item and a third two-item, within-category explicit order requirement condition.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the fifth of an incremental set of size-ordering phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the seventh ordinal position with the previous test condition now

serving as a control, the phases thereby providing two interspersed conditions of [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}B_{S3}C_{S2}$] and [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}B_{S3}C_{S1}C_{S2}$]. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the third category's smaller sized icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Only two monkey subjects had completed their eight-item fixed-ordering phases to date, achieving the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria were as shown in Table 5.7.

Table 5.7 Nos. of trials, errors and % errors to criterion [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}B_{S3}C_{S1}C_{S2}$]

Subject	No. of Trials	No. of Errors	% Errors
Mimi	300	202	67.3%
Luba	262	122	46.5%

Both subjects took fewer trials and errors to reach criterion than previously used to satisfy their [ABC] condition, and showed immediate successes, with 15% and 45% correct trials respectively amongst their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found for both subjects, for which very unequal representation of the possible error types was now evident. Accounting for between 79% and 90% of the total error trials completed for any one subject in this phase, it was found that despite there being forty-nine possible error paths executable, most could be accounted for by only six of them. For both monkeys the dominant error typologies reported were [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}B_{S3}C_{S2}$] (18%), [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S3}$] (16%), [$A_{S1}A_{S2}A_{S3}B_{S2}$] (12%) and [$A_{S1}A_{S3}$] (14%) all one-step within-category errors. The other two closely followed by [$A_{S1}A_{S2}B_{S1}$] (13%) and [$A_{S1}A_{S2}A_{S3}B_{S1}B_{S2}C_{S1}$] (13%), both one-step forwards-error category boundary errors. Both subjects showed some unusually large numbers of reiterative touches (10% of all errors), and, interestingly, nearly all of these were made in the

last (eighth) sequence position. For both subjects, RT analysis revealed significant rise times at the category boundaries only, although unequal variance did not allow data to be pooled. As shown in figure 5.7 below, the reaction time profile for Mimi reflected the category boundaries exactly, but with an extra pause being made before the last item to be touched ($p<0.05$ for all comparisons).

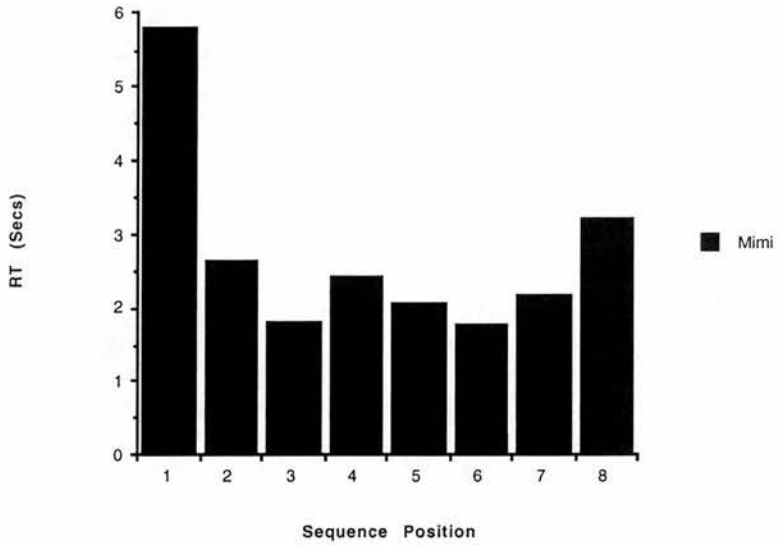


Figure 5.7 Mean RT profile for successful 8-item, three category fixed-order (size) sequencing by a Cebus monkey (Mimi)

For Luba, there was a weak category effect for the seventh position, but a significant increase occurred in the last position also ($p< 0.05$). Other pairwise comparisons showed that her RT profile otherwise reported (non-significantly) declining times to be taken for the first five post-IRT touches as she progressed through her eight-item sequence.

Discussion. This phase has produced the first clear evidence that the monkey would successfully report an eight-item series, and indeed here of a very special type: using three category markers in fixed relative sequence orders, comprising three exemplars each of the first two categories required to be reported in a fixed-order in the first six ordinal positions, to be followed by two ordered exemplars of a third class. Early successes were shown by both subjects (showing two and six correct respectively out of their first ten trials) and each completed the phase after fewer trials and errors than were previously required for them to complete phases of far less difficulty (e.g., [ABC]). Despite the possibility of

some forty-nine error types for the test condition in this phase, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 74% of the remaining error trials for which this strategy did not allow serial success. These latter error-types included both the familiar one-step forwards-error, (the subject moving on to start the next category before having touched all exemplars of the last) and the newer type of errors resulting from the novel requirement of fixed-order sequencing. These new error-types were also one-step forwards-errors, but this time took place *within* rather than across category boundaries, and accounted for more than a third of all errors recorded. Boundary rise time effects were revealed by analysis of individual RT data, and all were significant for Mimi, for each and every category. The phrasing here showed a return to the previously found tendency to 'chunk' the array into 'triplets', with Mimi showing a consistent three+three+one+one sequence execution, entirely consistent with the three categorical boundary positions for this sequence. This return to the more familiar chunking, consistent with the actual categorical boundary markers salient of the sequence, is probably due to the sequence now being of sufficient length as to be beyond the limits of the monkeys tolerable working memory for a simple linear serial-order production. Another subject, Luba, continued to show a systematic reduction in the time taken to touch each icon across her two serial monotonic sequences of the string, but then showed some weak rise time before starting the last category's exemplars (i.e., executing a chunking sequence of six+one+one). One may again continue to see here what amounts to clear evidence for generalisation and spontaneous class inclusion of novel items being added to existing sequences, mid-sequence, whilst demonstrating the existence and use of good strategic planning and economic management of increasingly larger fixed-order sequences of increasing complexity and level of difficulty.

Phase 31

Rationale. The successful completion of eight-item fixed-order sequences composed of three categories, each with multiple exemplars also to be explicitly ordered, was now further extendible, giving the monkey at this stage a final increase in string length by adding a further icon. With a view to determining whether the monkey would continue to

spontaneously classify and order similar icons in the array, this phase assessed the degree of generalisation and spontaneous classification in the serial order control of a novel nine-item set, again with a requirement for-category orderings, whilst simultaneously maintaining the existing categorical order as before. Increasing both in complexity and level of difficulty, it was again to be of interest to see whether, and in what ways, the category pausing would be effected in the presence of a simultaneous three-category, three-exemplar, explicitly-ordered nine-item sequence condition.

Subjects. Three adult monkeys (Charlie, Mimi and Luba) took part in this phase. (The first subject, Charlie, was seeing this nine-item condition having previously completed his six-item two-level hierarchical phase equivalent - see Phase 28).

Design and Procedure. This phase saw the sixth and last of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon into the ninth ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed strings of a control condition, $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}]$ and a $[A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}B_{s3}]$ test condition. An array example for this phase is shown in figure 5B.8. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the third category's largest sized icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this novel nine-item sequence seen as a fixed-order requirement condition. All subjects achieved the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control trials. The mean number of trials, errors and percentage errors to criteria were as shown in Table 5.8 below.

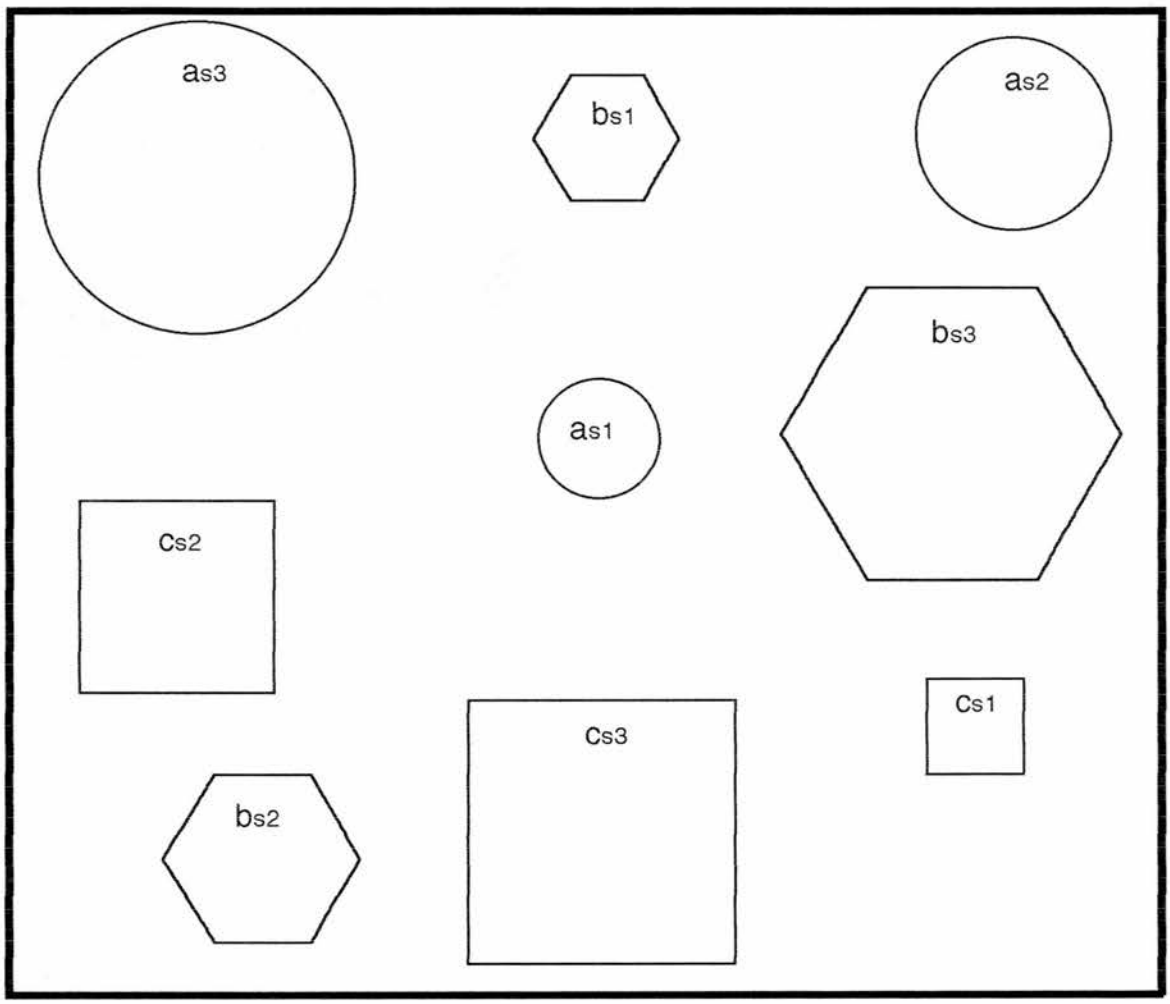


Figure 5B.7 Example of screen array for nine-item set with explicit size-ordering requirement (Phase 31)
[As1As2As3Bs1Bs2Bs3Cs1Cs2Cs3]

Table 5.8 Nos. of trials, errors and % errors to criterion
[$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}C_{s3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	1713	1031	60.2%
Mimi	178	105	59.0%
Luba	598	333	55.7%

All subjects except Charlie showed immediate successes with each showing between 30% and 40% correct trials amongst their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found across subjects, for which very unequal representation of the possible error types was still evident. Accounting for between 66% and 85% of the total error trials completed for any one subject in this phase, it was found that despite there being now sixty-four possible error paths executable, most could be accounted for by only seven of them. Charlie's [$A_{S}2$]-1st touch errors were still abnormally high (although never involving any other icon and making up 19% of all his error trials), and were due entirely to his initial carelessness in registration. For all monkeys the otherwise dominant error typologies reported were [$A_{S}1A_{s}2B_{s}2$] and [$A_{s}1A_{s}2A_{s}3B_{s}1B_{s}2C_{s}2$] (19%), both one-step forwards-category errors, followed by [$A_{S}1A_{s}3$], [$A_{S}1A_{s}2A_{s}3B_{s}2$], [$A_{S}1A_{s}2A_{s}3B_{s}1B_{s}3$] and [$A_{s}1A_{s}2A_{s}3B_{s}1B_{s}2B_{s}3C_{s}2$] (55%), all one-step within-category touch errors. Reiterative touches were now extremely low in number and would normally occur only in the last position. RT analysis revealed significant category boundary effects for only one of the three subjects, although many individual trials showed strong indications of pausing at both the first and second category boundary for all monkeys. The strongest effects for pausing were seen in the profile of Mimi's reaction times positions three, seven and nine ($p<0.05$), although these were not always coincident with the category boundaries (see fig 5.8). For Charlie and Luba, category boundary effects were either absent (in the case of the latter, who continued to show just a steady decline with a position nine peak) or not significant, as was the case for Charlie (see fig 5.9 below).

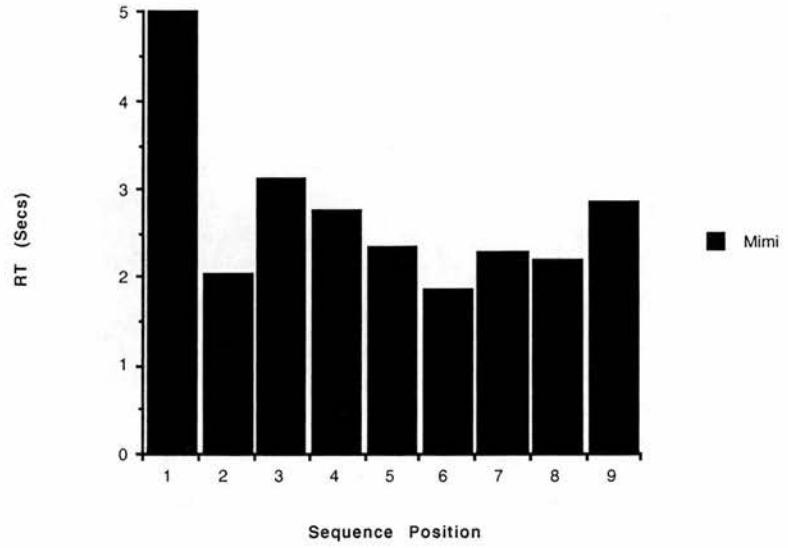


Figure 5.8 Mean RT profile for successful 9-item, three category fixed-order (size) sequencing by a Cebus monkey (Mimi)

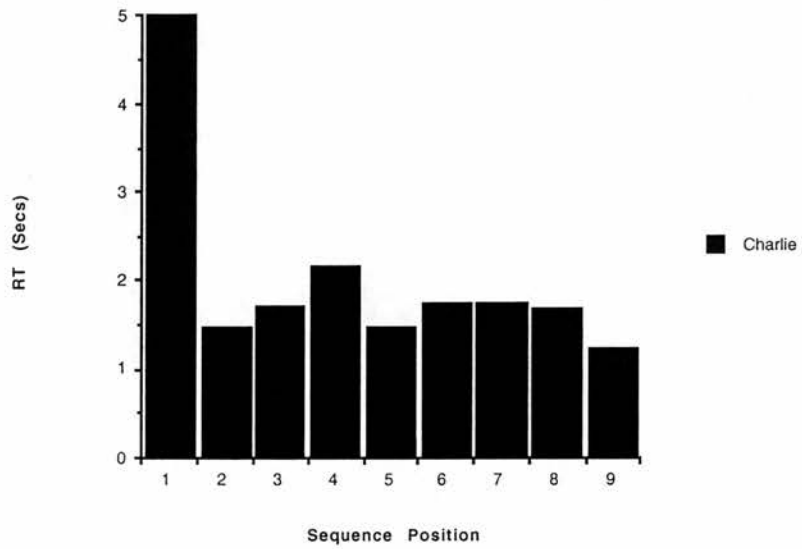


Figure 5.9 Mean RT profile for successful 9-item, three category fixed-order (size) sequencing by a Cebus monkey (Charlie)

Discussion (and interim summary of Phases 24-31). This phase has demonstrated for the very recordings of a non-human primate successfully, and consistently, reporting a nine-item series for which all items must be explicitly ordered, using three category markers in sequence order, each containing three exemplars. Each and every item was also to be reported in a fixed-order *within* each category and all three of the monkeys so far tested have succeded with this last phase of this

incremental series. Early successes were shown by all subjects except Charlie (each with two and four correct of the first ten trials respectively) and despite the increased difficulty, each completed the phase after taking fewer trials and errors than were previously required for them to complete their nine-item [ABC] sets. For example, in the case of Mimi, some 1950 trials (and 1659 errors) were initially required before she was to reach phase criterion for her first nine-item set as opposed to the 178 trials (and 105 errors) for this much more demanding two-level hierarchical condition of the same sequence-length. Despite there now being a total possible thirty-six error types for the test condition, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 85% of the remaining error trials for which this strategy did not allow serial success. These latter error-types were both the familiar one-step forwards-error (the subject moving on to start the next category before having touched all exemplars of the last), and the new one-step forwards-error seen within a category, now itself requiring its exemplars to be explicitly ordered. Category boundary effects were evident for selected individual trials, but were weak as revealed by more extended Bonferoni analyses of individual subject's RT data for criterion runs, for which only some were significant. The phrasing otherwise continued to vary idiosyncratically for each subject with Charlie showing a tendency to 'chunk' the nine-item array into a three+two+? item sequence execution, and Mimi preferring a two+four+? item chunking, the latter ignoring the more usual categorical phrasing pattern seen earlier. The remaining monkey (Luba) continued to show a systematic reduction in the time taken to touch subsequent icons as the sequence progressed. One can see here again that all of the monkeys so far tested have demonstrated clear generalisation and transfer of their previously learned sequencing skills and have provided good evidence for the existence and use of some strategic planning with increasingly larger sequences.

No monkey has so far failed to seriate a single condition of this latter series to date and each monkey appears to execute two-level hierarchical seriation in a way which did not express itself in any unprincipled, stereotyped fashion when classification options were available under supervised conditions (see summary table 5.9 below).

Table 5.9 Numbers of monkey subjects completing Phases 25-31 [ABC] 2-level hierarchy conditions (N=100%)

Phase No.	Fixed order test condition	No. of monkeys to criterion
25	$A_{s2}B_{s2}C_{s2}$	3
26	$A_{s1}A_{s2}B_{s2}C_{s2}$	3
27	$A_{s1}A_{s2}A_{s3}B_{s2}C_{s2}$	3
28	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}C_{s2}$	4
29	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s2}$	3
30	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}$	2 *
31	$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}C_{s3}$	3

* A third monkey (Alfie) was working on this condition at the time of writing

Although incidental to these quite momentous acquisition achievements, after the string-length had been extended beyond four to five items, there was a re-emergence of chunking paterns becoming apparent once more, although less pronounced (see tables 5.10a-b) . Furthermore, because of the more salient size-seriation cues now made available in this series of experiments, this provided further evidence that the monkey might be able to take advantage of such privileged salience (relational monotonicity) in order to reduce the memory load required for locating the current sequence position on-line.

Table 5.10a Monkey subjects showing significant latency rise-times only at the category boundaries 2-level (size) hierarchy phases.

Subject	4-items		5-items		6-items	
	1st	2nd	1st	2nd	1st	2nd
Charlie	-		-		◊	◊
Alfie	x	x	x	x	◊	x
Mimi	x	x	x	◊	x	◊
Luba	x	x	x	x	x	x

Table 5.10b Monkey subjects showing significant latency rise-times only at the category boundaries 2-level (size) hierarchy phases.

Subject	7-items		8-items		9-items	
	1st	2nd	1st	2nd	1st	2nd
Charlie	-		-		x	x *+6*
Alfie	x	x *+6*	-		-	
Mimi	x	◊ *+5*	◊	◊ *+8*	x	◊ *+3,9*
Luba	x	x	◊	x *+8*	x	x

◊= ($p<0.05$); x = NS; *+n,..*= other sig. ($p<0.05$) RT position (s).

It was most noticeable that the monkeys these later chunking preferences were not always consistent with the category boundary markers, yet always supporting a high degree of successful seriation. If such an effect was indeed the result of the nature of the particular stimuli employed in this last 2-level hierarchical seriation task (for which a highly salient, non-arbitrary, monotonic size order requirement was imposed upon the subject) one might not expect to see such variation in phrasing for conditions employing equally salient, yet more arbitrary connectives. The following series of experiments were designed to address this issue using the same procedure, again requiring explicit fixed-ordering of each and every item of the array, affording categorical seriation in tasks employing a more arbitrary *within*-category sequence of otherwise equivalent levels of difficulty.

(b) 2-level categorical (colour) seriation - acquisition phases

Introduction

Based this time upon a more arbitrary colour coding sequences, the conditions of the second group of two-level hierarchical seriation experiments, Phases 32 through 39, were as shown in Table 5.11 below and include the details of interspersed control conditions and completion criteria for each phase. All of the three monkeys participating in this experimental series started with Phase 32 and thereafter, the monkeys followed the subsequent phase steps in the order listed now that the sequence of increments for their previous size-ordering phases had been established.

The following series of phases would require the monkey to attempt yet another two-level hierarchical series, but this time using an explicit colour-ordering rule within each of the same three shape categories up to and including a nine-item sequence, following the same pattern as for the two-level size-ordering phases above. The three colours used for each category were the same as those employed in the original [ABC] set for [A], [B] and [C] respectively in Phase 3, but this time were used for the colours of each category alike (e.g., $[A_{c1}]$, $[A_{c2}]$ and $[A_{c3}]$).

Table 5.11 Serial order Phases 32-39 [ABC]
2-level hierarchy: 3 colours/3 categories

Phase No.	Fixed order Conditions	Completion criterion
32	$A_{c1}A_{c2}A_{c3}$ $B_{c1}B_{c2}B_{c3}$ $C_{c1}C_{c2}C_{c3}$	15/20 15/20 15/20
33	$A_{c1}B_{c2}C_{c3}$	15/20
34	$A_{c1}B_{c2}C_{c3}$ $A_{c1}A_{c2}B_{c2}C_{c3}$	3/4 15/20
35	$A_{c1}A_{c2}B_{c2}C_{c3}$ $A_{c1}A_{c2}A_{c3}B_{c2}C_{c3}$	3/4 15/20
36	$A_{c1}A_{c2}A_{c3}B_{c2}C_{c3}$ $A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c3}$	3/4 15/20
37	$A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c3}$ $A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c3}$	3/4 15/20
38	$A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c3}$ $A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}$	3/4 15/20
39	$A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}$ $A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c2}C_{c3}$	3/4 15/20

All of these new phases required continuously high levels of vigilance and consistent performance in the face of increasing task difficulty. Task difficulty here was not only increasing with respect to the combinatorial explosion of possible pathways through increasing search spaces with larger string-lengths (see fig. 3.2), but were further to be exacerbated by there now being a far greater numbers of error-types for any given string-length than before, now that the explicit ordering of every icon of the set be required. Another feature of this incremental series was that the novel icon to be added when increasing string-length would be inserted at different points of the sequence dependant upon the nature of the string composition most recently completed. Thus a new string-length would not simply result from the addition of an extra icon to the end of the existing string. Furthermore, the rise in the level of difficulty with increasing string-length could also be greatly enhanced as a result of the high degree of potential colour conflict offered by the screen arrays, for which a number of exemplars of different categories might share the same colour characteristic. One possible solution to a merely exhaustive search of such an array might be to classify them on the basis of colour alone at the expense of shape, but in order to demonstrate the existence of a more controlled two-level hierarchical structure as outlined in figure 3.3 (and for comparison with the size-order conditions), the shape

categories were to serve as the 'superordinate' class for subdivision under the present supervised learning conditions. For all subsequent phases, individual monkeys were self-selecting by their continuing successful attainment of the previously given phase criteria as they progressed through tasks of increasing string length and greater levels of difficulty. The actual sequence compositions presented during these colour-ordering acquisition phases were drawn from a larger number of possible decomposition strings derivable from an [ABC] icon set, but were of a type more familiar to each subject and those conceivably simpler in structure by comparison.

For every phase condition in this next experimental series, a unique category and category exemplar order was required to be reported by each subject for each icon set in the array of every trial. Thus, for successful completion of all the coming phases, every new task demanded an explicit and unique ordering of each and every individually identifiable icon upon the screen array. From trial to trial the same icons were presented in different configurations so as to avoid the production of stereotyped motor responses, whilst at the same time provoking continuous array interrogation on the part of the subject. The use of an interspersed control condition (the test condition of the previous string length and compositionality) was also given partly for this reason, so preventing the subject's responses becoming too reflexive and thereby less 'cognitive' in production, whilst simultaneously allowing interpretation of any significant amounts of error in the face of greatly increasing task demands under supervised learning conditions.

Phase 32

Rationale. To allow the demonstration of the existence of a base-line from which any depth of colour-based search might be deemed supportable, this first of the 2-level (colour) hierarchy phases was given so as to inform us that the pre-existing requirement for the subject being able to order the individual icons of each category be attainable independantly. This first phase would also inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli in a consistent and orderly manner, simultaneously for each class.

Subjects. Three adult monkeys (Charlie, Mimi and Luba) have to date taken part in this phase.

Design and Procedure. In accordance with the general procedure above, Phase 32 presented three conditions [$A_{c1}A_{c2}A_{c3}$], [$B_{c1}B_{c2}B_{c3}$] and [$C_{c1}C_{c2}C_{c3}$], each of which required that the three icons of each category [A], [B] or [C] (using the same three colours for each category) were to be reported in the order, say, red star- green star- blue star. Each category's exemplars were to be interrogated and touched sequentially - [X_{c1}] first, followed by [X_{c2}], followed by [X_{c3}] (where X=category marker) - thus requiring of the subject a unique solution to be found for every array presented. Four error-types were possible (three forwards errors and one reiteration), repeated touches to the last touched icon still being permissible. The individual conditions were given in a random order over 50 trials on a ratio schedule of 1:1:1. Phase criteria required the subject to complete fifteen correct trials out of the last (cumulative) twenty for each condition. Upon reaching a condition criterion, that particular condition would not be presented again during the phase.

Results. All three subjects completed this three-condition phase successfully and were eligible for continuation with the two-level (colour) hierarchical phases of the experimental series. The mean number of trials, errors and percentage errors to criteria for the conditions are shown in Tables 5.12-14.

Table 5.12 Nos. of trials, errors and % errors to criterion [$A_{c1}A_{c2}A_{c3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	248	154	62.0%
Mimi	242	148	61.1%
Luba	266	145	54.5%

Table 5.13 Nos. of trials, errors and % errors to criterion [$B_{c1}B_{c2}B_{c3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	104	60	57.7%
Mimi	165	109	66.0%
Luba	317	209	65.9%

Table 5.14 Nos. of trials, errors and % errors to criterion [$C_{c1}C_{c2}C_{c3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	31	15	48.3%
Mimi	190	122	64.2%
Luba	496	289	58.2%

All subjects showed immediate success with between 10%-70% correct trials out of their first twenty for any one condition, with order of completion varying amongst subjects. Analysis of the incorrect trials revealed a consistent pattern across monkeys, and typically more than 55% could be accounted for by a single one-step forwards-error touch for all conditions, although there were four possible error paths executable. RT analysis revealed significantly decreasing time to be taken to touch the subsequent icons of every condition for all subjects ($p<0.05$), and with equal variance allowing data pooling, figure 5.10 shows a similar profile to that of previous three-item phases.

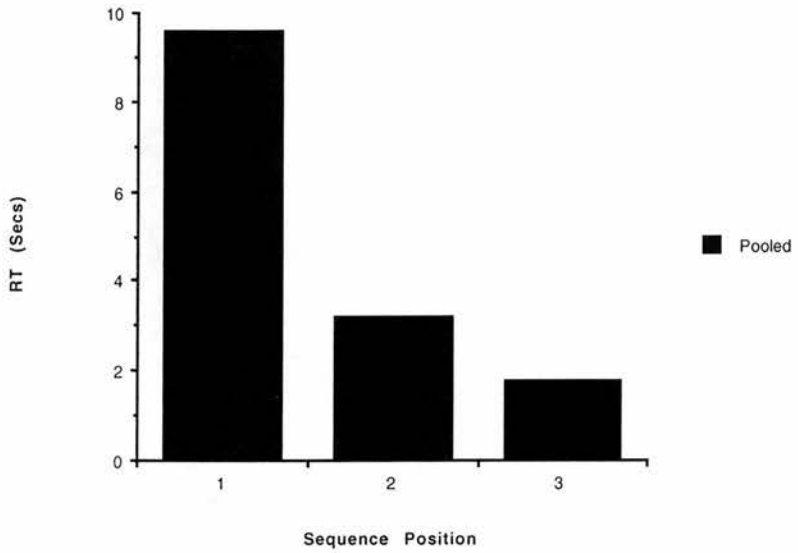


Figure 5.10 Mean RT profile for successful 3-item, three category fixed-order (colour) sequencing by a Cebus monkeys (N=3, Pooled)

Discussion. It has now been established that each monkey would consistently report each of the category's three-item colour sequences independently with the same explicit ordering requirement being asked of each class. All conditions excepting Luba's [B] and [C] sets were completed with less time and effort (as measured by trials and errors to

criteria) than that required to reach the same level of performance with their initial [ABC] three-item set. Further, for the pooled data, most of the errors recorded were of one type, [$X_{c}2$], suggestive of good generalisation of previous ordering patterns to these novel fixed-order sequences, such preservations offering neither evidence for sequential search ambiguity or other ordering difficulties. Confirming this trend, the RT findings of all three monkeys show that the time taken to search each subsequent icon of the set to be ordered decreased throughout sequence execution, suggestive of the development of a route-planning strategy to be taking place. The initial reaction times (IRTs) were unusually large here compared to those for recent performances with size-ordering strings of much greater length, reflecting greater search time to be initially needed at the outset of these novel stimuli sets.

Phase 33

Rationale. To permit the firm establishment of the new 'core' sequence for subsequent generalisation and extension, a single test condition [$A_{c}1B_{c}2C_{c}3$] using the same three previously seen categories, (but now again with the full colours and order of the original [ABC] set) were presented for discrimination and ordering. This phase would again inform us as to the reliability of each monkey's ability to differentiate between its particular stimuli in a robust, and fixed-ordered manner.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. In accordance with the general procedure as before, this phase presented a single condition [$A_{c}1B_{c}2C_{c}3$] with all subjects preserving their three [ABC] category icons, discriminable by shape and spatial location shown simultaneously to the touch-screen. Although now carrying a different nomenclature, this particular phase presents an identical set of arrays to those of he original [ABC] icon set. These icons were to be interrogated and touched sequentially - [$A_{c}1$] first, followed by [$B_{c}2$], then finally [$C_{c}3$], requiring of the subject a unique solution to be found. Four error-types were possible (three forwards errors and one reiteration), with repeated touches to the last touched icon still being permissible. The subject had completed this phase by achieving fifteen correct out of the last (cumulative) twenty trials.

Results. All three monkeyss completed this condition and thereby qualified for continuation to the next phase of their two-level hierarchical series of arrays. The mean number of trials, errors and percentage errors to criterion for the condition is shown in Table 5.15.

Table 5.15 Nos. of trials, errors and % errors to criterion
$[A_{c1}B_{c2}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	26	7	26.9%
Mimi	18	3	16.6%
Luba	32	12	37.5%

All three Cebus completed this new 'core' sequence acquisition condition almost immediately demonstrating good maintenance of this prior learned sequence, last seen some eight months previously. All subjects showed some error, mostly of one type (a touch to $[B_{c2}]$ first). Reaction time (RT) profiles for the last fifteen correct trials (those for which the most consistent and reliable data may be obtained) indicated significantly decreasing times to be taken by all of the monkeys to touch subsequent icons once they had begun to interrogate the touch-screen array ($p<0.05$). Between-subject variation was very low, and pooled RT mean data for the six animals was as shown in Figure 5.11.

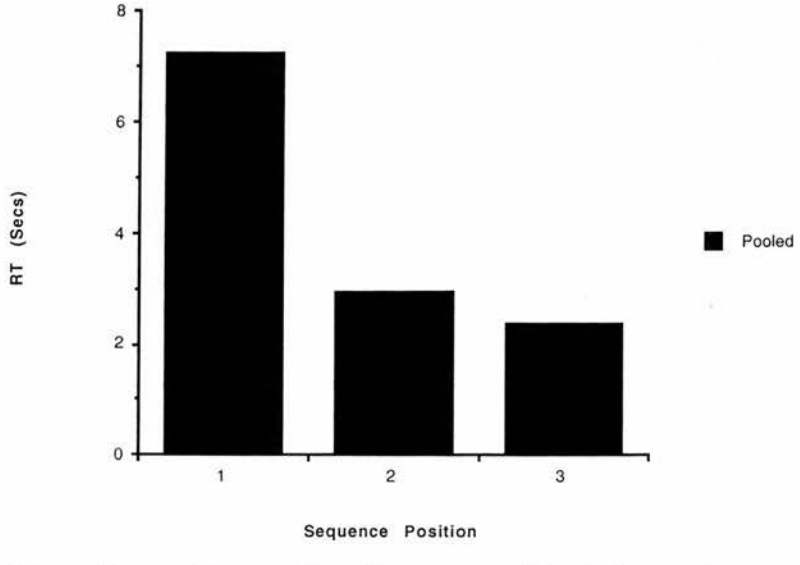


Figure 5.11 Mean RT profile for successful 3-item, three category fixed-order (colour) sequencing by a Cebus monkeys (N=3, Pooled)

Discussion. It has quickly been established that each monkey would consistently report their 'core' colour-ordered, three-item sequences following a supervised training procedure which allowed no flexibility of ordering on the part of the subject. Essentially a rerun of the earlier [ABC] training phase, it can be seen that 'new' three item sequence was almost immediately recognised. For all subjects the only errors recorded were of one type, [$B_{c}2$] - first, with closer inspection of Charlie's video-cassette tape revealing poor screen-touching (and hence registration) to account for much of them. The immediate high levels of success and the absence of three out of the four possible error types was suggestive of all monkeys demonstrating high degrees of conservation of class ordering from their earlier training phases. Support for this claim derived from the RT findings of all three monkeys, for which the time taken to search each subsequent icon of the set to be ordered decreased throughout sequence execution, suggestive of the development of a route-planning strategy to be taking place during the initial time to first touch.

Phase 34

Rationale. With a view to determining whether the monkeys were going to be capable of continued spontaneous classification of known familiar icons in an array, the next phase set out to explore the possibility of generalising their serial order control to a novel four-item set, having increased the sequence length by one icon, but this time adding the need for explicit *within*-category colour orderings as well as maintenance of the existing class order. This novel core-sequence expansion would not only offer an indication of the monkey's ability to spontaneously classify a principally orderable set, but would also, if successful, provide us with the first examples of a non-human primate to be capable of working with a colour-based hierarchical grouping principle, albeit under supervised learning conditions. Indications of the monkey's having derived any early success as a result of the use of data reducing strategies would be clear from accelerating acquisition rates following increases in sequence length and difficulty that these next phases provide.

Subjects. Two adult monkeys (Mimi and Luba) took part in this phase (a third monkey, Charlie, had previously moved on directly to what had by now become Phase 35 below).

Design and Procedure. In the same way as for the size-ordering series, this phase saw the first of an incremental set of phases for which the previous test condition was to be raised in string length by one icon in order to extend the sequence to be searched in order to produce a nine-item string with three-categories, each to contain three orderable exemplars within them. This first colour-condition increment sees the addition of an icon to the second ordinal position and thereby presents two interspersed conditions of $[A_{c1}B_{c2}C_{c3}]$ and $[A_{c1}A_{c2}B_{c2}C_{c3}]$. The two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the first category's second coloured icon (now the same colour as that of the second class exemplar) to the test condition of the previous phase. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Both subjects completed this phase, achieving the 75% criterion level required, whilst simultaneously maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.16.

Table 5.16 Nos. of trials, errors and % errors to criterion
$[A_{c1}A_{c2}B_{c2}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Mimi	16	1	6.0%
Luba	17	2	12.0%

Both of the subjects showed immediate successes with 88% and 97% correct trials for Mimi and Luba respectively being seen within their first twenty. Analysis of the incorrect trials revealed a consistent pattern to be found, with both subjects only showing one or two single-step forwards errors in the first position, despite there being nine possible error paths executable. RT analysis revealed significant category boundary effects only for one subject, the phrasing effect shown by Mimi for the third and fourth positions ($p< 0.05$) giving a profile as seen in figure 5.12 below consistent with the category boundaries of this condition. The reaction

time profile of Luba was to show a steady (and non-significant) reduction in the times taken for subsequent touches throughout her sequence production.

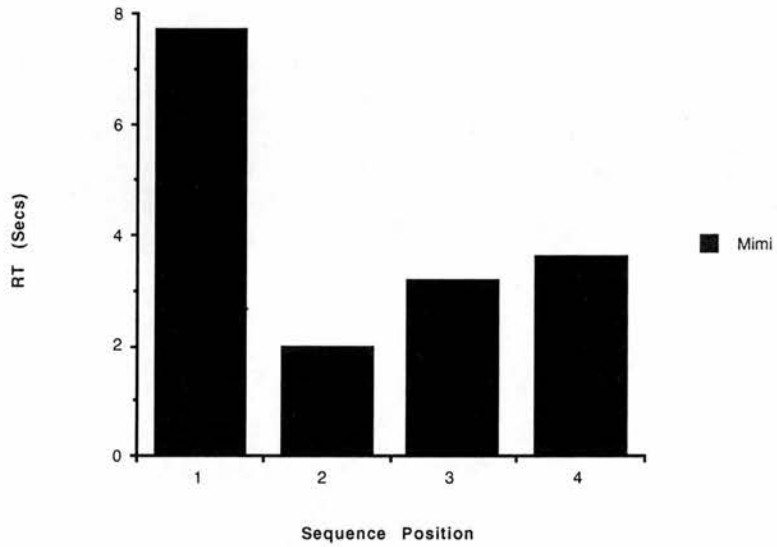


Figure 5.12 Mean RT profile for successful 4-item, three category fixed-order (colour) sequencing by a Cebus monkey (Mimi)

Discussion. The criteria performance measures have shown that both of the monkeys successfully reported a further four-item set with novel compositionality, using all three category markers with each maintaining their relative sequence order with two fixed-order coloured exemplars in the first two ordinal positions. Early successes were shown by both subjects (each with nine correct of the first ten trials) and despite the increased level of difficulty, each completed the phase after taking fewer trials and errors than were required for them to complete their previous consolidating three-item sets. Despite the possibility of some nine error types for the test condition now, error profiles revealed clear evidence of the monkeys having 'seen' the solution right from trial one, with only one or two forwards errors making up the entire failure space. Boundary effects were revealed by analysis of only one of the subject's RT data, the preference of the other being for consistently phrasing a single four-item chunk suggestive of little extra cognitive load resulting for her successful execution such a small sequence. This was so despite the requirement to ignore the conflicting colour cues for the purposes of within-class ordering.

Phase 35

Rationale. Having satisfactorily demonstrated the ability to negotiate a four-item fixed-order coloured sequence composed of three categories, the first with two exemplars, it was now possible to give the monkey another increase in string length following the addition of a further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order information to a novel five-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and level of difficulty, indications of the monkey's having derived any early success as a result of the use of data reducing strategies would be clear from accelerating acquisition rates and the development of any boundary effects for individual subject RT profiles. It was to be of particular interest to note whether category boundary effects would emerge in the presence of this three-item within category order requirement condition.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the second of this incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the third ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{c1}A_{c2}B_{c2}C_{c3}]$ and $[A_{c1}A_{c2}A_{c3}B_{c2}C_{c3}]$. This new icon bore the same colour as that of the last category's exemplar, there now being two icons of the [A] class which were differentiable from both [B] and [C] exemplars only by shape. The two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the first category's third coloured icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this phase, achieving the 75% criterion level required for the test condition, whilst simultaneously maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.17.

Table 5.17 Nos. of trials, errors and % errors to criterion
$[A_{c1}A_{c2}A_{c3}B_{c2}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	19	8	42.1%
Mimi	67	34	50,7%
Luba	217	131	60.3%

All three subjects showed immediate success on trial one, going on to demonstrate between 30% and 45% correct trials out of their first twenty. Analysis of incorrect trials revealed a consistent pattern across subjects, in which unequal representation of the eighteen possible error typologies was evident. For as much as 80% of the total error trials completed for any one subject in this phase, it was found that the same two error paths could account for the incorrect touches made. The dominant error type reported was $[A_{c1}A_{c2}B_{c2}]$ (55%), a one-step category boundary forwards-error, followed by $[A_{c1}A_{c3}]$ (30%) a one-step within-category forwards-error. RT analysis revealed only one significant category boundary effect, which for case of Mimi, showed a marked rise in the search time for the first and second category ($p< 0.05$) as may be seen from figure 5.13. For both Charlie and Luba, steadily declining reaction times were shown with each subsequent touch in a manner similar to that seen in previous phases (see figure 5.14). No other pairwise comparisons were found to be significant.

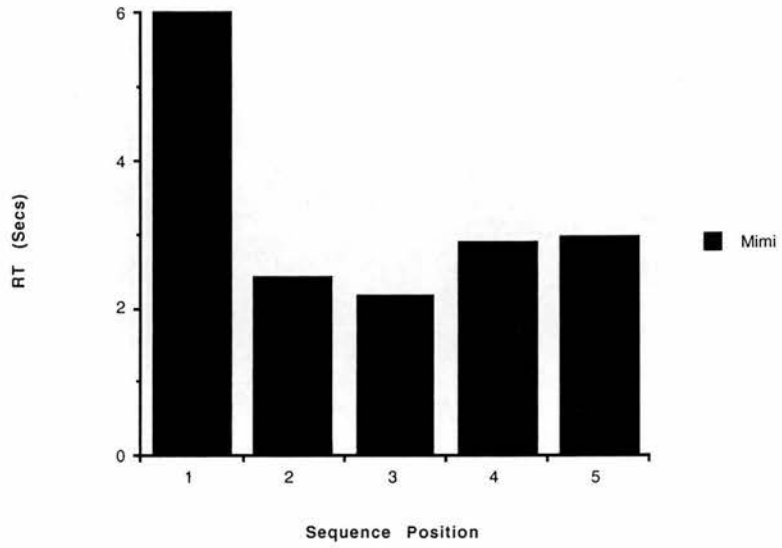


Figure 5.13 Mean RT profile for successful 5-item, three category fixed-order (colour) sequencing by a Cebus monkey (Mimi)

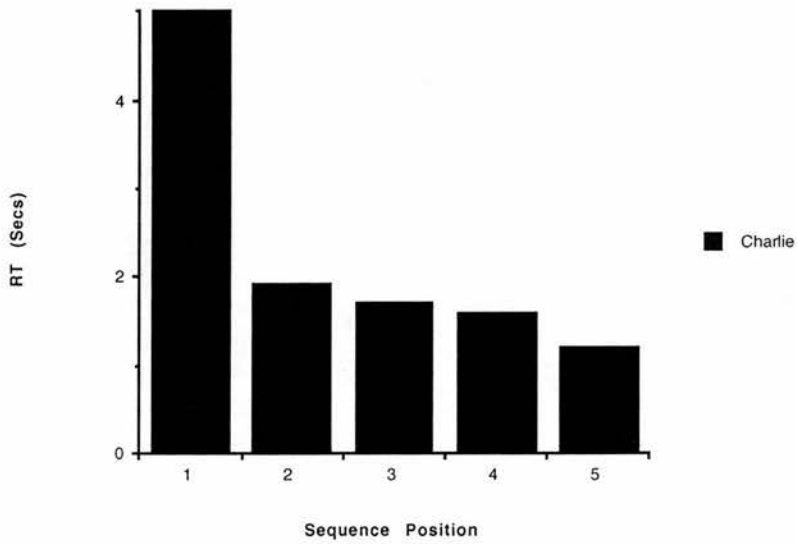


Figure 5.14 Mean RT profile for successful 5-item, three category fixed-order (colour) sequencing by a Cebus monkey (Charlie)

Discussion. This phase demonstrates once more that the monkey would successfully report a five-item series, using three category markers maintain in a relative sequence order, together with three exemplars to be reported in a fixed-order in the first three ordinal positions. Early successes were shown by all subjects (each with between five and nine correct of the first ten trials) and despite the increased difficulty, each

completed the phase after taking markedly fewer trials and errors than were previously required for them to complete their last five-item sets (cf. Phase 27). Despite the possibility of some eighteen error types for this test condition, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 80% of the remaining error trials for which this strategy did not allow serial success. These latter error-types were both familiar one-step forwards-errors, the subject moving on to start the next category before having touched all exemplars of the last. Significant category boundary effects were revealed by analysis of RT data for Mimi who showed a tendency for three+two chunking. For the other two subjects the phrasing appeared to be produced as a single five-item chunk, the RT profile showing systematic but non-significant reductions in the time taken to touch subsequent icons as the sequence progressed. We do see here, again, that all monkeys are demonstrating clear generalisation and transfer of their previously learned sequencing skills and good evidence for the existence and use of some strategic planning (or at least economic management) with a novel larger sequence when classification options were presented under supervised conditions.

Phase 36

Rationale. Having satisfactorily demonstrated the ability to negotiate a second five-item fixed-order sequence composed of three categories, the first with three exemplars, it was now conceivable to give the monkey another increase in string length following the addition of one further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order control to a novel six-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and level of difficulty, it was to be of interest to see in what ways the category boundary effects might alter in the presence of one three-item and a second two-item within-category order requirement condition.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the third of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the fourth ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{c1}A_{c2}A_{c3}B_{c2}C_{c3}]$ and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c3}]$. The two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the second category's first coloured icon. This new icon bore the same colour as the first of the initial category but was to be grouped by shape in the second class. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed their six-item fixed-ordering phases, achieving the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.18.

Table 5.18 Nos. of trials, errors and % errors to criterion $[A_{c1}A_{c2}A_{s3}B_{c1}B_{c2}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	355	179	50.4%
Mimi	109	47	43.1%
Luba	126	80	63.5%

All three subjects showed immediate successes with between 15% and 50% correct trials among their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found across subjects, for which very unequal representation of the possible error typologies was now evident. Accounting for between 69% and 78% of the total error trials completed for any one subject in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only three of them. Although Charlie's error profile continued to show many $[A_{c2}]$-1st touch errors (20%), for all monkeys the dominant error typologies reported were $[A_{c1}A_{c2}A_{c3}B_{c2}]$ (30%) and $[A_{c1}A_{c2}B_{c1}]$ (7%), both one-step forwards-category errors, followed by

[$A_{S1}A_{S3}$] (20%) a one-step within-category touch error. All subjects recorded very few reiterative touches. RT analysis revealed significant rise times consistent with the category boundaries only for both female subjects ($p<0.05$), with equal variance allowing data to be pooled, as shown in figure 5.15 below. For Charlie, there was consistent (but non-significant) phrasing at the third and fifth positions with no main effect of category noticeable at all.

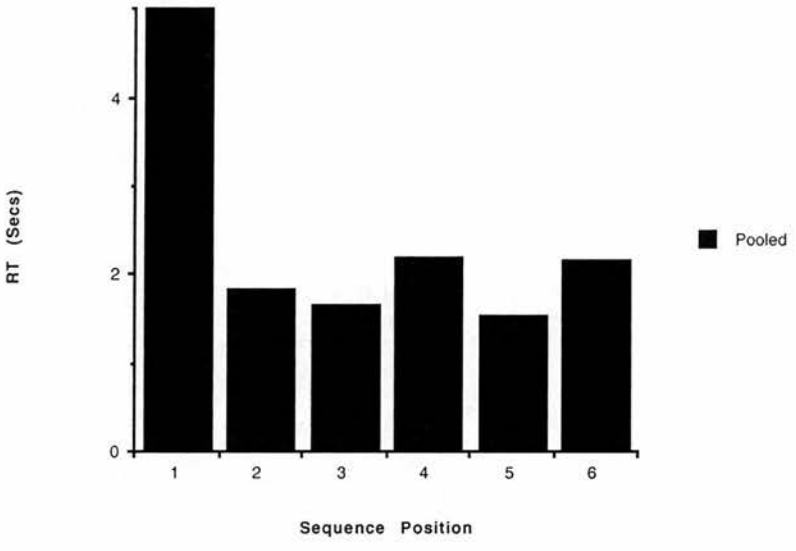


Figure 5.15 Mean RT profile for successful 6-item, three category fixed-order (colour) sequencing by Cebus monkeys (N=2, pooled)

Discussion. This phase has produced clear evidence that the monkey would successfully report a six-item series, using three category markers in a fixed relative sequence order, with two or three exemplars of the first two categories required to be reported in a fixed-order in the first five ordinal positions. Early successes were shown by all subjects (showing between two and five correct of the first ten trials) and each completed the phase after fewer trials and errors than were previously required for them to complete phases of equivalent levels of difficulty. Despite the possibility of some twenty-five error types for the test condition in this phase, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 76% of the remaining error trials for which this strategy did not allow serial success. These latter error-types included both the familiar one-step forwards-error, (the subject moving on to start the next category before having touched all exemplars of the last) and the new type of errors

resulting from the novel requirement of fixed-order sequencing. These new error-typologies were also one-step forwards-errors, but this time took place *within* rather than across category boundaries, and accounted for around a third of all the errors recorded. Although clear boundary effects were revealed by analysis of individual and pooled RT data, they were significant for only two of the monkeys. The phrasing for another subject, Charlie, showed a consistent but non-significant tendency to 'chunk' the array into a two+two+two sequence execution, entirely dissonant with the three categorical boundary positions for this sequence. Charlie was perhaps able to 'see at a one glance' what the solution could be, and planned the entire six touches to be made at the outset, although given the particular icon colours contained in this array, one might have better expected to see the emergence of a three+two+ chunking being favoured. This re-emergence of phrasing consistent with the category boundaries for two of the monkeys are perhaps indicative of the subjects responding to the increasing memorial load being placed upon them for the execution of the longer six-item string-length. There does, nonetheless, continue to be clear evidence for the generalisation and spontaneous class inclusion of the novel items being added to existing sequences, mid-sequence, whilst consistently demonstrating the existence and use of some strategic planning and economic management of increasingly larger fixed-order sequences of increasing complexity and difficulty.

Phase 37

Rationale. The successful completion of six-item fixed-order sequences composed of three categories, the first and second with multiple orderable exemplars, is now further extendible, giving the monkey yet another increase in string length by adding one further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase assessed the degree of possible generalisation of serial order control with a novel seven-item set, again with a requirement for explicit*within*-category orderings, whilst simultaneously maintaining the existing category order. Increasing both in complexity and level of difficulty, it was again to be of interest to see in what ways the category boundary effects would be effected in the presence of a simultaneous three-category, three- or two-

exemplar, within-category order requirement in the same condition.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the fourth of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon into the sixth ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c3}]$ and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c3}]$. The degree of colour replication was now quite large: each of the first two categories contained an exemplar of the same colour, the third colour being represented in all three classes. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the second category's third coloured icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. All three monkeys completed this novel seven-item sequence seen as a fixed-order requirement condition. All subjects achieved the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control trials. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.19.

Table 5.19 Nos. of trials, errors and % errors to criterion
$[A_{c1}A_{c2}A_{s3}B_{c1}B_{c2}B_{c3}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	108	62	57.4%
Mimi	142	78	54.9%
Luba	470	287	61.0%

Subjects showed immediate success, with all showing at least 60% correct trials in their first twenty. Analysis of the incorrect trials revealed quite a consistent pattern to be found across subjects, for which very unequal representation of the possible error typologies was still evident. Accounting for between 74% and 94% of the total error trials completed for any one subject in this phase, it was found that despite there being

twenty-five possible error paths executable, most could be accounted for by only four of them. For all monkeys the dominant error typologies reported were [$A_{c1}A_{c2}B_{c2}$] (30%), the usual one-step forwards-category error, followed by [$A_{c1}A_{c2}A_{c3}B_{c1}B_{c3}$] (22%), [$A_{c1}A_{c3}$] (17%) and [$A_{c1}A_{c2}A_{c3}B_{c2}$] (15%), the one-step within-category touch errors. All subjects showed only a few error trials involving any reiterative touches. RT analysis revealed significant rise times at the first and second boundaries only ($p< 0.05$), although for only two monkeys were they significant for the second ($p< 0.05$). The profile of pooled reaction times were as shown in figure 5.16 below.

Discussion. This phase has provided a second demonstration of the monkey successfully reporting a seven-item series, using three category markers maintained in relative sequence order, together with the three exemplars of the first two categories being reported in fixed-orders in the first six ordinal positions. Early successes were shown by all subjects (each with between one and five correct of their first ten trials) and, despite the increased difficulty, each completed the phase after taking fewer trials and errors than were previously required for them to complete their original nine-item sets.

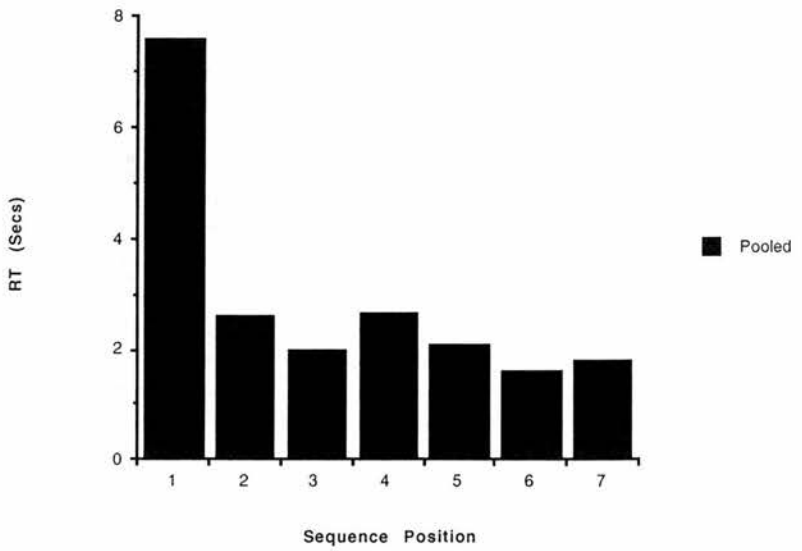


Figure 5.16 Mean RT profile for successful 7-item, three category fixed-order (colour) sequencing by Cebus monkeys (N=3, pooled)

First touch errors were recorded in fewer trials than ever before, as were reiterative touches. Despite the possibility of some thirty-six error types

for the test condition now, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for as much as 79% of the remaining error trials for which this strategy did not allow serial success. These latter error-types were both the familiar one-step forwards-error (the subject moving on to start the next category before having touched all exemplars of the last), and the new one-step forwards-error seen within a category, now itself requiring its exemplars to be explicitly ordered. Category boundary effects were evident for all subjects, and were all significantly so bar one monkey for the second phrase marker. The phrasing otherwise showed no idiosyncratic variations for any subject, with each showing a tendency to 'chunk' the seven-item array into a three+three+one item sequence execution, a return to the more usual categorical phrasing pattern seen earlier with sequences of this length. Once again, all monkeys continue to demonstrate clear generalisation and transfer of their previously learned sequencing skills and provide good evidence for the existence and use of some strategic planning with increasingly larger sequences. It therefore continues to be of interest to see in what ways this might continue for an increasingly larger search space.

Phase 38

Rationale. Having again satisfactorily demonstrated the ability to negotiate a seven-item fixed-order sequence composed of three categories, the first two each with three differently coloured exemplars, it was now reasonable to give the monkey a further increase in string length following the addition of one new icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase set out to explore the generalisation of serial order control to a novel eight-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining the existing categorical order. The complete array now contains two categories with the three same colour orders in each, and for the last category, the two colours are the same as those of the first two ordered positions of the other two classes. Increasing the search set both in complexity and level of difficulty, it was to be of interest to see in what ways the category boundary effects would be consolidated in the presence of two three-item and a third two-item, within-category explicit order

requirement condition.

Subjects. Three adult monkeys took part in this phase.

Design and Procedure. This phase saw the fifth of an incremental set of colour-ordering phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon to the seventh ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed conditions of $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c3}]$ and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}]$. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the third category's third coloured icon. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Only two subjects have completed their eight-item fixed-ordering phases to date, achieving the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control condition. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.20.

Table 5.20 Nos. of trials, errors and % errors to criterion
$[A_{c1}A_{c2}A_{s3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	1020	588	57.6%
Mimi	410	209	50.9%
Luba	-	-	-

Both of the subjects took slightly more trials and errors to reach criterion than previously used to satisfy their equivalent size-ordering condition, but did show some early successes, with 35% and 40% correct trials for Charlie and Mimi respectively, amongst their first twenty. Analysis of the incorrect trials revealed the usual consistent pattern to be found for both subjects, for which very unequal representation of the possible error typologies remained evident. Accounting for some 81% of the total error trials completed for either subject in this phase, it was found that despite

there being forty-nine possible error paths executable, most could be accounted for by only seven of them. For both monkeys the dominant error typologies reported were $[A_{c1}A_{c3}]$ (16%), $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c3}]$ (16%), $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c2}]$ (12%) and $[A_{c1}A_{c2}A_{c3}B_{c2}]$ (10%) all one-step within-category forwards-errors. The other three were $[A_{c1}A_{c2}B_{c1}]$ (15%), $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c1}]$ (10%) and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c2}]$ (10%), the latter group all one-step forwards-errors at the category boundaries. Both subjects showed some unusually large numbers of reiterative touches (<10% of all errors), and interestingly nearly all of these were made in the last (eighth) sequence position. RT analysis revealed significant rise times to occur only at the category boundaries only ($p<0.05$) for both subjects. With equal variance allowing data to be pooled, Figure 5.17 below shows the reaction time profile for both monkeys to reflect the first and second category boundaries exactly ($p<0.005$) . Other post-hoc evaluations revealed either non-significant differences or a significantly decreasing RT in the case of the third sequence position ($p<0.05$).

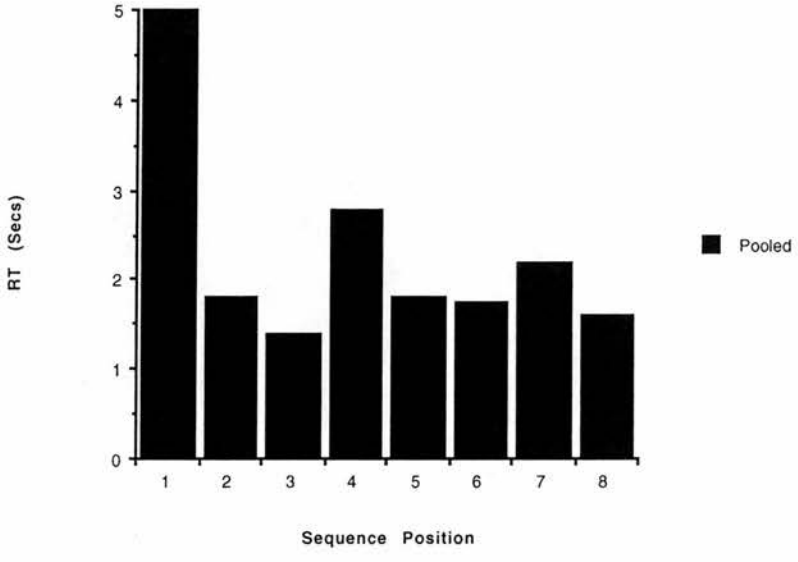


Figure 5.17 Mean RT profile for successful 8-item, three category fixed-order (colour) sequencing by Cebus monkeys (N=2, pooled)

Discussion. This phase has produced a second demonstration of the monkey successfully reporting an eight-item series, and again, here of a very special type: three category markers were to be used in a fixed relative sequence order, comprising three exemplars each of the first two categories, themselves required to be reported in a fixed-order in the first

six ordinal positions, to be followed by two ordered exemplars of a third class. Early successes were shown by both subjects, each showing three and four correct out of their first ten trials. Despite the possibility of some forty-nine error types for the test condition in this phase, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 80% of the remaining error trials for which this strategy did not allow serial success. These latter error-types included both the familiar one-step forwards-error, (the subject moving on to start the next category before having touched all exemplars of the last) and the newer one-step forwards-errors, this time taking place *within* rather than across category boundaries. Boundary effects were revealed by analysis of pooled and individual monkey RT data, and were found to be significant for both at the first and second category boundaries. The phrasing here shows a return to the previously found tendency to 'chunk' the array into 'triplets', with both subjects showing a reliable three+three+two sequence execution, entirely consistent with the categorical boundary positions for this sequence. This return to the more familiar chunking consistent with the actual categorical boundary markers, is probably due to the sequence now being of sufficient length as to be beyond the limits of the monkeys tolerable working memory for a simple linear serial-order production. We again continue to see here what amounts to clear evidence for generalisation and spontaneous class inclusion of novel items being added to existing sequences, mid-sequence, whilst demonstrating the existence and use of good strategic planning and economic management of increasingly larger fixed-order sequences of increasing complexity and level of difficulty.

Phase 39

Rationale. The successful completion of the second eight-item fixed-order sequences composed of three categories, each with multiple exemplars to be explicitly ordered, is now further extendible, giving the monkey at this stage a final increase in string length by adding a further icon. With a view to determining whether the monkey would continue to spontaneously classify and order similar icons in the array, this phase assessed the degree of generalisation and spontaneous classification in the serial order control of a novel nine-item set, again with a requirement for explicit *within*-category orderings, whilst simultaneously maintaining

the existing categorical order as before. Increasing both in complexity and level of difficulty, it was again to be of interest to see whether, and in what ways, the category pausing would be maintained in the presence of a simultaneous three-category, three-exemplar, explicitly-ordered nine-item sequence condition.

Subjects. Two adult monkeys (Charlie and Mimi) took part in this phase.

Design and Procedure. This phase saw the sixth and last of an incremental set of phases for which the previous test condition was to be raised in string length by one icon. This novel sequence saw the addition of an icon into the eighth ordinal position with the previous test condition now serving as a control, the phases thereby providing two interspersed strings of a control condition, $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}]$ and a $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c2}B_{c3}]$ test condition. An array example for this phase is shown in figure 5B.7. For all subjects the two conditions were presented at a 1:4 ratio for control and test conditions, the latter having the addition of the third category's second coloured icon. The appearance of the screen array was such that all the three categories contained one each of the same three colours for explicit ordering within each class. After each and every twenty test trials, maintenance levels of three out of four successful control trials were required before the test condition continued to be presented. The phase was completed when the test condition criterion of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Both monkeys completed this novel nine-item sequence seen as a fixed-order requirement condition. All subjects achieved the 75% criterion level required for the test condition, whilst maintaining a high standard of performance with the control trials. The mean number of trials, errors and percentage errors to criteria are shown in Table 5.21.

Table 5.21 Nos. of trials, errors and % errors to criterion
$[A_{c1}A_{c2}A_{s3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c2}C_{c3}]$

Subject	No. of Trials	No. of Errors	% Errors
Charlie	494	292	59.1%
Mimi	519	302	58.1%

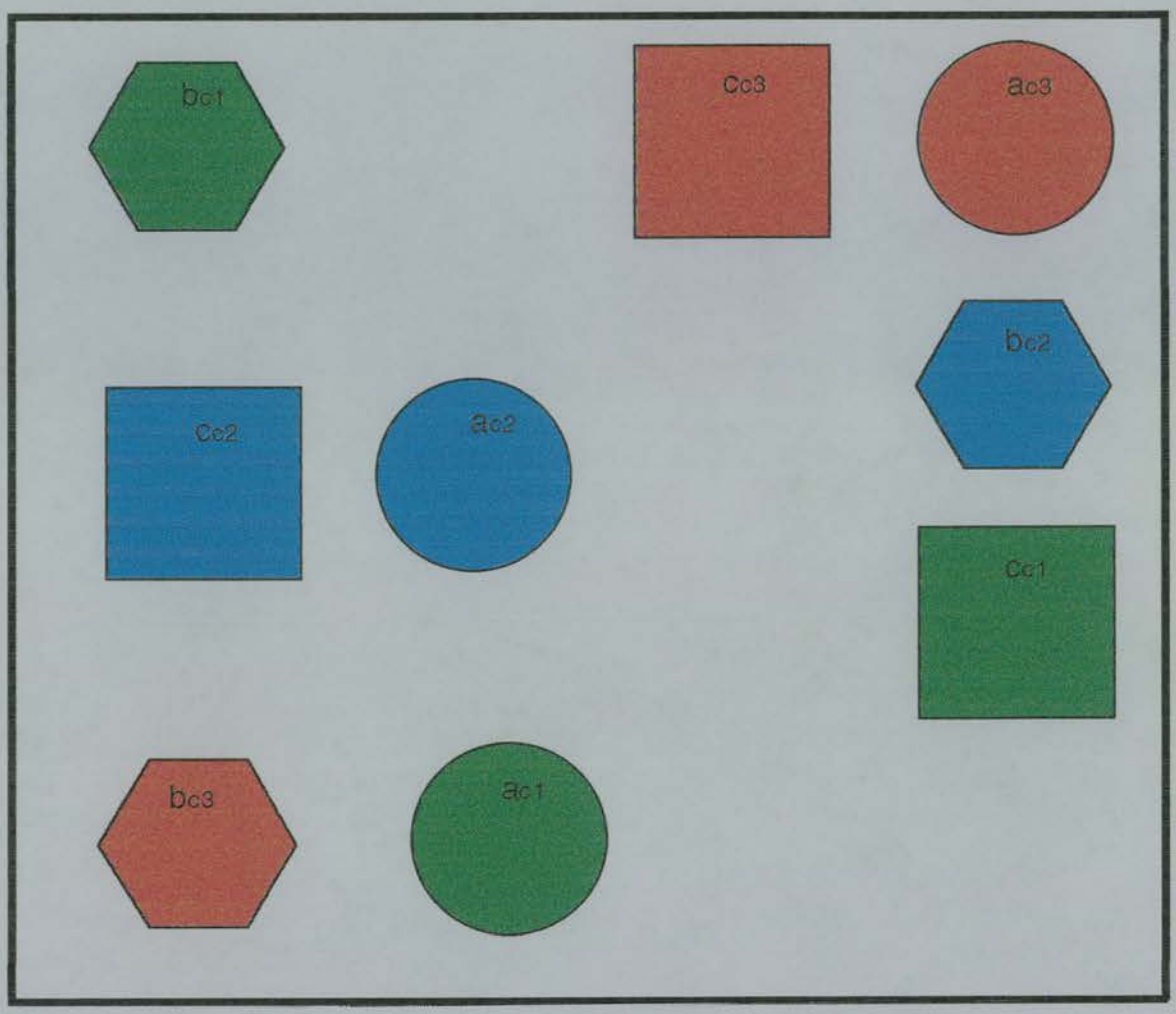


Figure 5B.8 Example of screen array for nine-item set with explicit colour-ordering requirement (Phase 39)
[$A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c2}C_{c3}$]

Each subject showed immediate successes with 10% and 35% correct trials amongst their first twenty for Charlie and Mimi respectively. Analysis of the incorrect trials again revealed a consistent pattern to be found for both subjects, for which very unequal representation of the possible error typologies was still evident. Accounting for between 65% and 72% of the total error trials completed for any one subject in this phase, it was found that despite there being now sixty-four possible error paths executable, most could be accounted for by only six of them. Charlie's $[A_{c2}]$-1st touch errors were still abnormally high (although never involving any other icon and making up 6% of all his error trials), and were due entirely to his initial carelessness in registration. The dominant error typologies reported were $[A_{c1}A_{c2}B_{c2}]$ and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}C_{c2}]$ (26%), both one-step forwards-category errors, followed by $[A_{c1}A_{c3}]$, $[A_{c1}A_{c2}A_{c3}B_{c2}]$, $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c3}]$ and $[A_{c1}A_{c2}A_{c3}B_{c1}B_{c2}B_{c3}C_{c1}C_{c3}]$ (45%), all one-step within-category touch errors. Reiterative touches were now extremely low in number and for both monkeys would normally occur only in the last position. RT analysis revealed significant rise times again to occur at the category boundaries ($p<0.05$) for both subjects equal variance allowing pooling to produce the profile as shown in figure 5.18. Other post-hoc evaluative measure for pairwise comparisons revealed non-significant differences between intra-categorical touches except for the reduced RT for position three as seen in the previous phase.

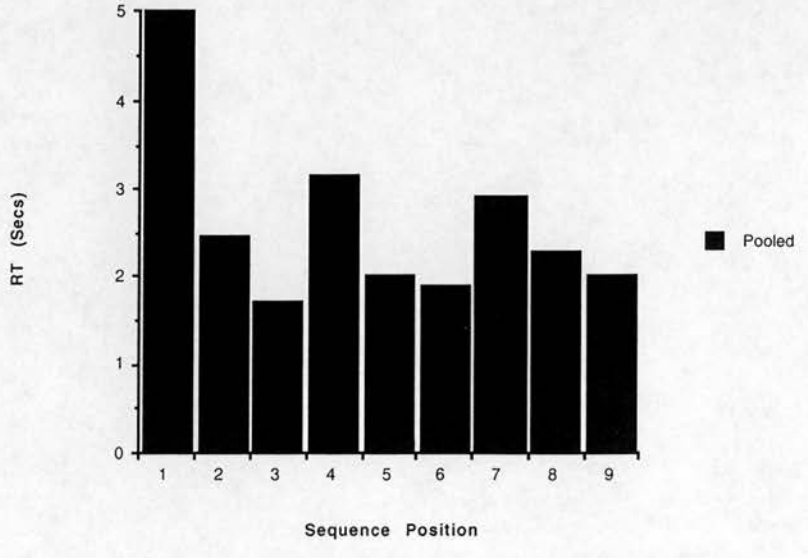


Figure 5.18 Mean RT profile for successful 9-item, three category fixed-order (colour) sequencing by Cebus monkeys (N=2, pooled)

Discussion. This phase has demonstrated for a second icon feature that the monkey would successfully report a nine-item series, for which all items must be explicitly ordered, using three category markers in sequence order, each containing three exemplars, themselves to be reported in a fixed-order within each category. Early successes were shown by both subjects (Charlie with one and Mimi with two correct of their first ten trials) and despite the increased difficulty, each completed the phase after taking far fewer trials and errors than were previously required for them to complete their original nine-item [ABC] sets. Despite the possibility of some sixty-four error types for the test condition now, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 70% of the remaining error trials for which this strategy did not allow serial success. These latter error-types were both the familiar one-step forwards-error (the subject moving on to start the next category before having touched all exemplars of the last), and the new one-step forwards-error seen within a category, now itself requiring its exemplars to be explicitly ordered. Category boundary effects were evident for both subjects, showing the most consistent and significant phrasing to date for any RT data analysed for criterion runs. The phrasing showed unambiguous 'chunking' of the nine-item array into a three+three+three arbitrary colour sequence execution. This latter phase completes the two-level (colour) hierarchy serial order phases, resulting in a subject successfully interrogating and reporting a second nine-item sequence of individually identifiable icons both for three categories, and this time simultaneously ordering by colour *within* each category. To date, all monkeys so far tested have consistently demonstrated clear evidence for the generalisation and transfer of their previously learned sequencing skills whilst provide good indicants of the existence and use of strategic planning with increasingly larger and more complex sequences.

Summary of results for two-level hierarchical organisation phases 24-39.

The results of this second set of experiments have provided the first documented evidence for the existence of two-level hierarchical organisation and control of serial fixed-order production in a non-human primate. Successful executions of sequences with up to, and including,

nine items have been robustly produced for the first time by the monkey for which explicit ordering requirements were in place for each and every item of the array. This first implementation of a second stage in the behaviour-based paradigm of McGonigle and Chalmers (1993, 1996) continues to support their view that a cognitively economic agent would be able to demonstrate the emergence of data-reducing strategies such as classification and 'chunking' in a serial order task of sufficient length and complexity. Over the last eighteen months, four *apella* monkeys have to date been taken through a series of supervised learning experiments which provoke a continued need to adapt their previously acquired abilities to control serial order productions of three-category sequences. Such adaptive strategies have been continualy developed by the monkey in order that they become able to cope with novel sequence strings for which the exemplar features of each category might vary, but remain to some degree equivalent for the purposes of sequential analysis and production. Although by no means a small achievement in itself, further extensions of these two-level hierarchical tasks in terms of breadth (merely prolonging the sequences in a linear fashion) would have eventually imposed severe memory constraints upon the subject, the end result at the very best being to place an upper limit figure for each monkey's manageable sequence lengths. The aim, however, of this second set of experiments was to determine whether, and to what extent, the monkey might be able to report the ability to maintain their current levels of performance whilst also being required to explicitly order the items within each category. Such an ability has allowed the monkey to demonstrate the emergence of a two-level hierarchical organisation of array items, in which a within-category requirement was presented with familiar categories of exemplar icons for which a superordinate categorical classification option remained open.

At present, there are six subjects working through various stages of this second part of the study, and to date, no monkey subject has failed to complete any given phase or condition. This is also no meagre achievement, and as never before seen with so many subjects in a ongoing series of experiments of this kind, the continued success of every subject despite raising the stakes in terms of levels of difficulty are even greater here than they were for the previous phases in the earlier part of the study reported above. Not only was there still the effect of a combinatorial

explosion to be managed on-line (the number of possible pathways through the search space increasing with increasing string length), but for the explicit ordering of a given sequence's every exemplar, there was also the larger exponential rise in the number of error-types with increasing string length as illustrated in figure 5.19 below). By chance factors alone, one might expect both the numbers and types of errors made by the monkey to increase with the production of ever longer sequences, but their continuing performances suggest the emergence of a high degree of economic, managerial control via the use of generalisable strategies based upon classification.

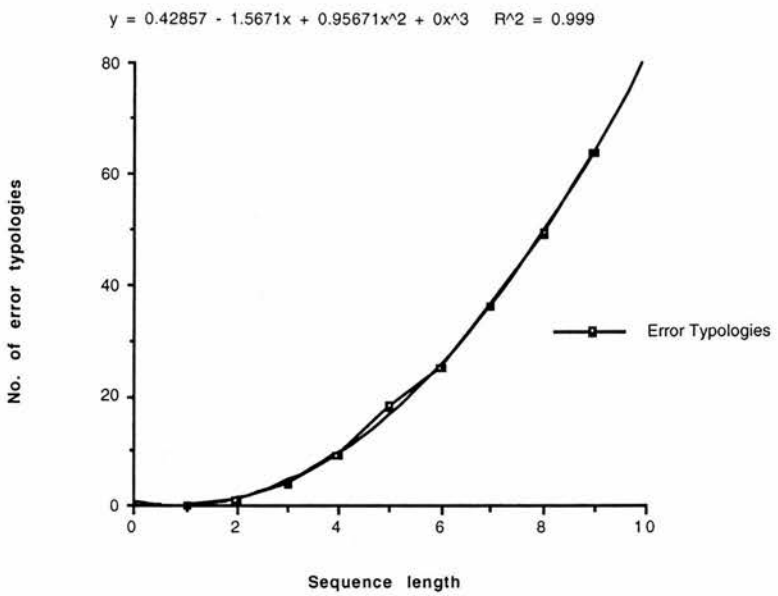


Figure 5.19 Exponential rise in error-space with increasing set size.

Indeed, as may be seen from figure 5.20 below, for acquisition measures of a the same monkey's sequences for increasing complexity and levels of difficulty, there is no evidence for the otherwise expected exponential rises in effort being required to reach successive phase criteria. Good evidence for generalisation effects were also to be found in the high incidence recording of spontaneous class inclusion following the presentation of each new array. Furthermore, for every new phase, because the condition length increases involved the insertion of new icons to a different sequence position each time, no simultaneous chaining methodology could have been employed in coming to a ready solution. This makes it even more significant the finding that there were at least two correctly completed trials within the first ten of each subsequent first

session for each phase.

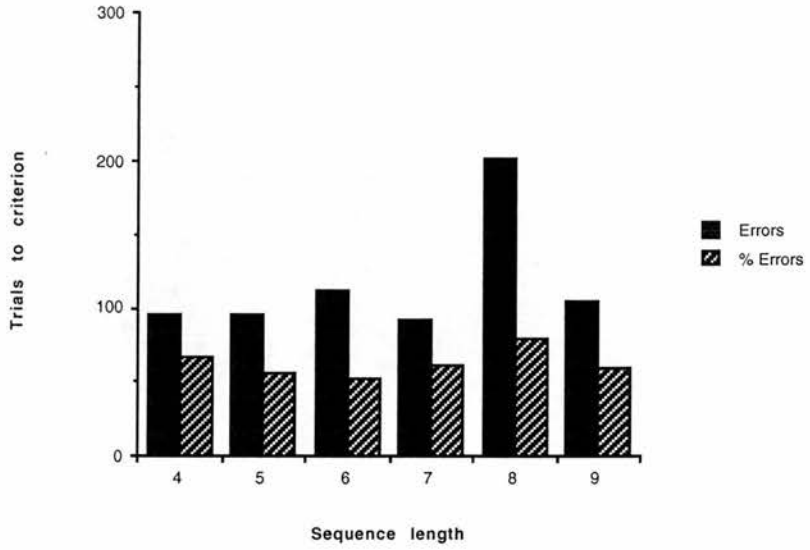


Figure 5.20 Numbers of errors and %errors to reach criterion for increasing set size in a Cebus monkey (Mimi).

Such increases in the task demands with increasing set sizes have also led the subject to generate production strategies which reached beyond the requirements of the task given, some of which may be determined here from a review of time data analysis. At a first glance, it might be thought that the phrasing effects produced as a result of sequencing would continue to reflect chunking in accordance with the changing memorial demands of sequence length, but on closer inspection, there were more subtle and idiosyncratic effects to be noticed dependant upon both subject and the sequence compositionality under interrogation. Chunking preferences were relatively consistent for each monkey, but for the explicit size-ordering series of experiments did not always provide clear patterns consistent with the category boundaries (cf table 5.10a-b above). Given the added salience of an intrinsic orderable sequencing based upon relational size differences alone, and given that the within-category exemplar number was at most three items, the subjects appeared capable of producing longer chunks of four or sometimes five touches before pausing. Indeed, phrasing structures of four+one, or five+one item sequencing became quite common following expertise, often to penetrate across the previously binding effects of category boundary. One interesting exception consistently found in the size-ordering series, was that of Charlie who did not undergo the incremental procedure, but was

given only [$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}$] and [$A_{s1}A_{s2}A_{s3}B_{s1}B_{s2}B_{s3}C_{s1}C_{s2}C_{s3}$] phases. He was the only subject to show a consistent three+three and three+three+three chunking preference for these conditions. For the colour-ordering phases, as may be seen in table 5.22a-b below, all subjects were found to phrase in accordance with the class boundaries once beyond the six-item stage, perhaps suggestive of an adaptive strategy preventing the effects of an otherwise salient grouping of the icons by colour categories, a method not allowed by condition programming constraints.

Table 5.22a Monkey subjects showing significant latency rise-times only at the category boundaries 2-level (colour) hierarchy phases.

Subject	4-items		5-items		6-items	
	1st	2nd	1st	2nd	1st	2nd
Charlie	-		x	x	x	x
Mimi	◊	◊	◊	x	◊	◊
Luba	x	x	x	x	◊	◊

Table 5.22b Monkey subjects showing significant latency rise-times only at the category boundaries 2-level (colour) hierarchy phases.

Subject	7-items		8-items		9-items	
	1st	2nd	1st	2nd	1st	2nd
Charlie	◊	x	◊	◊	◊	◊
Mimi	◊	◊	◊	◊	◊	◊
Luba	◊	◊	-		-	

◊= ($p<0.05$); x = NS; $+n,..$= other sig. ($p<0.05$) RT position (s).

It was noteworthy in support of this idea, that one monkey (Charlie) did indeed appear to be attempting to solve his six-item series in this way at first, initially recording a number of [AD] errors as if in an attempt to collapse the set into a 'first-things-first' principled search as seen in the combined [ABCDEF] condition (cf: Phase 22). This major difference also reflects an important distinction between the relational properties of the arbitrary and non-arbitrary kinds of stimuli sets presented for the two-level hierarchical classification series. The greater salience afforded by the non-arbitrary monotonic size-relational features led the subject to rely upon the first-level categorical distinctions to keep track of their sequence position, an option not so easily determined when seriating the arbitrary colour-based relational connectives.

Given that we are now in a better position to assess the levels and degrees

of any spontaneous classificatory competence of the developing monkey subject, any future explorations of the continuing growth of cognitive processes using this technique could be built upon the existing exhibited and known of the now expert subject. The monkeys participating in this ongoing, longitudinal series of experiments are showing good evidence to suggest that they are substantially benefiting from the cumulative knowledge gained from their dynamic, interactive, and life-historical evaluation of task successes following the acquisition, transfer and execution of two-level hierarchical search expertise as reported above.

Overview and conclusions of experimental series (Phases 1-39) to date

The experimental design and procedures which have been implemented and described in this thesis have given rise to the very first unambiguous demonstrations of the control of serial order productions of up to twelve-item sequences and provided clear evidence for a two-level hierarchical organisation in the non-human primate (*Cebus apella*). From the outset of this open-ended 'voyage of discovery', there was never any guarantee that the monkey would succeed with the levels of task complexity that have now been reported above and at best one might have expected to have reproduced the findings of D'Amato and Columbo (1988), who had previously recorded a maximum five-item series, also with an *apella* monkey. Although not every subject has been presented every condition set once beyond the acquisition phases 1-11, no individual monkey has been removed from the experimental series due to failure with any of the increasingly difficult tasks. All of the original cohort of six adult *apella* subjects continue to present daily for transfer to the testing room and willingly engage with the laboratory apparatus. An unprecedented 100% continuity and success with a monkey species, atypical of that reported in the primate learning literature, Table 5.23 below reviews the current landmark completions for each monkey to date, the falling numbers reflective of the current phase engaged by an individual in the series, not their failure to complete the increasingly straining task demands.

Table 5.23 Numbers of monkey subjects completed Phases 1-39 to date (N=6=100%)

Phase Nos	Test conditions	No. of monkeys to criterion
1-9	Acquisition [ABC]: 3-10-item set	6
11	Acquisition [ABC]: 12-item set	3
12-13	Colour equivalence sets [ABC]: - 9-items only	5
14	Size equivalence sets [ABC]: - 9-items only	4
15-17	Acquisition [DEF]: 3-9-item set	3
18-19	Colour equivalence sets [DEF]: - 9-items only	3
20	Size equivalence sets [DEF]: - 9-items only	3
21-22	Free Search [ABCDEF] 6-items only	2
24-31	Two-level (size) hierarchy: 1-9 item sets	3
32-39	Two-level (colour) hierarchy: 1-9 item sets	2

All of the six monkey subjects participating in this study over the last four years have demonstrated successful seriations with the expansion of at least one 'core' three-icon sequence to produce a number of nine-item sequences composed of three categories, each with three exemplars of each category. Further, whether these exemplars were physically equivalent or free to vary in colour or size features, offering a free choice within a category, the icons nonetheless continued to be classified using the invariant superordinate criteria of category shape as the basis for class inclusion. Three of the monkeys were later exposed to set sizes of as many as twelve items, and each subject proceeded to successfully complete that phase without the suggestion of their having reached an upper limit of their series expansion along that dimension (one level down from a simple linear string of increasing numbers of new categorical exemplars). Rather than going on to discover what such upper limits might actually have been for each monkey, a second series of experiments exploring a possible extension by 'depth' of search was undertaken with the knowledge that extended classifications of the previous kind would have to break down after reaching some sequence length beyond which the individual category sizes were themselves so large as to start imposing search problems of their own. This second level of search was to have provided a solution to this latter problem if managed in the form of a

nested hierarchical organisation, thereby reducing the immediate search-space to a more restricted sub-set of the overall array.

Given that it had already been established that the monkeys could work with category exemplar variances in both colour and size, the subjects were thereafter required to explicitly order a nine-item set, but this time also for the icons within each category whilst maintaining the original class orderings as before. For these latter two-level hierarchical phases of the experimental series, all six adult monkeys are now progressing through various stages of both colour- and size-ordering conditions and none have to date failed to reach criterion levels of performance with any of the tasks so far undertaken of this type. Three monkeys have successfully been taken forward to an ordered nine-item set with three sizes of each of three categories, and two of them have reached the same stage with a three-colour version of the same task.

Consistent with claims made for the emergence of data-reducing strategies on the part of both the human and monkey subject, a micro-analysis of the characteristic touching patterns of the monkey was undertaken in the search for any evidence for the hypothesis of McGonigle and Chalmers (1992) which suggested that in the face of increasingly complex and difficult tasks given over a protracted period of time, a subject would compensate by the production of progressively economic, organisational structures. Following the growth towards expertise of each individual monkey throughout their series of serial search tasks as reported above, indicants of a progressive adaptation to tasks of increasing levels of difficulty have continued to provoke the emergence of competences suggestive of a cognitive organisation previously unseen (or at least underdetermined) in traditional learning experiments in the laboratory.

A first indicant of the monkey's progressive adaptation to tasks of increasing levels of difficulty was derived from measures of the acquisition functions for the [ABC]-core sequence and its subsequent expansion to twelve-item string-lengths. Given the combinatorially expansive set of possible pathways through the prospective search space with increasing set size (see figure 3.2) it was very clear from all of the landmark summaries given above, that for all subjects, there was no equivalent exponential increase in either the numbers of trials or errors

to reach criterion for successively longer sequences. Indeed, for two out the three monkeys succeeding with the twelve-item sets, fewer trials were required than for their previous nine-item phases. Such results suggest that the monkeys are sensitive to the use of classificatory structures for the purpose of data-reduction of an otherwise too larger set for serial control without the use of brute-force memorial effort. Together, the successive generalisations based upon classification and transfer measures, have been consistently found with the production of extended sequences and are indeed seen to have significant consequences for the solving of the much more complex problems of the later tasks.

A further indicant of this economic advantage of classification may be seen from an analysis of the reaction time profiles for successful sequencing by each individual monkey. Although incidental to the acquisition discoveries to be assessed in these experiments (measures of latency were neither an explicit experimental variable, nor were they manipulated or controlled for in any way) it has been consistently found in the course of this ongoing series of phases, that with sequences of differing lengths and compositionalities, a number of RT profiles were to emerge which showed clear phrasing effects, often coinciding with the category boundaries inherent in a given sequence. Such a chunking effect was similar to that reported by McGonigle and Jaswal (1993) in which spontaneous classification was seen to be produced by young nursery children with the same task under comparable conditions. In the case of this human-child study, the subjects ably demonstrated an [ABC]-core sequence extended to include a fifteen-item sequence (five categories with three identical exemplars of each), but also showed a consistent RT phrasing pattern related to the category boundaries (see fig. 3.4 above). Furthermore, not only did the individual latencies rise at every category boundary during sequence execution, as the subject progressed down the sequence, the time taken to touch the first icon of each new category became shorter although there were no significant changes in the intra-category latencies throughout the sequential production. This finding may be interpreted to suggest that the subject was not just making pauses, but that once the first item had been identified, then the subsequent items to be searched required relatively less effort as the sequencing continued. Latency effects for the monkeys in this study have not proven to be so clear (although neither were the same longer sequence compositions

attempted) but a similar profile was produced for nine-item seriation for both the monkey and the young human subject. Furthermore, if the first touch latency was to be taken as reflecting a measure of the overall search time of the array, it was certainly worthy of note here that the initial reaction times were seen to increase with the later two-level hierarchical condition phases. As may be seen from table 5.24, these phrasing effects produced by the monkey are not merely acquisition phenomena (compare also figures 5.10a-b and figs 5.22a-b). Taking the mean RT values for criterion run trials, the various nine-item sequence profiles were largely consistent, although idiosyncratic variations did arise. For all subjects the initial nine-item set gave rise to highly significant phrasing effects at both category boundaries and have essentially been maintained throughout, although with varying degrees of significance.

Condition	Charlie	Mimi	Alfie	Luba	Ollie	Kissy
AAABBBCCC	⇑* ⇑*	⇑* ⇑*	⇑* ⇑*	⇑* ⇑*	⇑* ⇑*	⇑* ⇑*
Monochrome	⇑* ⇓	⇑* ⇑*	⇑* ⇑*	<=> ⇑*	⇑* ⇑*	-
Colour Equiv.	⇑* ⇑	⇑* ⇑*	⇑ ⇑	⇑ ⇑	-	-
Size Equiv.	⇑ ⇑	⇓* ⇑*	⇑* ⇓*	-	⇑* ⇑*	-
DDDEEEFFF	⇑ ⇑	⇑ ⇑*	⇑ ⇑*	-	-	-
Colour Eqiv.	⇑ ⇓*	⇑ ⇑	⇑ ⇑*	-	-	-
Size Equiv.	⇑ ⇑	⇑* ⇑*	-	-	-	-
Size-ordering	⇑ <=>	⇓ ⇑*	-	⇓ ⇓	-	-
Colour-order	⇑* ⇑*	⇑* ⇑*	-	-	-	-

Key: ⇑= increase RT; ⇓=decrease RT; <=> = no change; *= sig. $p < 0.05$.

Table 5.24 Direction and significance of first and second category boundary effects for various nine-item sequences by Cebus monkeys

A further indication of the significance of the monkey's phrasing effects was to be seen in cases for which clearly differential RT profiles were produced according to the compositionality of sets with identical sequence lengths. For example, with both the mixed six- and ten-item conditions (see results and discusions of Phases 6 and 9) all pairwise comparisons revealed significant rise times to occur *only* at the categorical boundaries, despite these boundary locations varying between conditions for the same subject within the same session. If such phrasing effects were to be determined by

sequence length *per se*, one would not have expected there to be such consistent and spontaneously principled RT variation occuring on-line according to the compositionality of the set to be seriated. Such a finding is certainly not an artifact of the task demands and is indicative of the emergence of a clearly useful strategy linked to high levels of performance and seriation expertise. On occasion (and particularly in the case of size variation conditions) the boundary RT change did not so much produced a pause, but an acceleration in the time to touch the first icon of the following category. Such salience as might be attached to a monotonic-size series appears to have allowed the subject to transcend the category boundary for as long as the number of icons in the chunks as actually parsed, remain relatively low. Such an affordance was not made an option in the colour-ordering conditions for which there were both arbitrary connectives to be learned and colour conflicts for class inclusion across the categories to be disambiguated. These differences were perhaps made more evident by the fact that each monkey made a return to the production of consistent and significant phrasings coincident with both the category boundaries whilst working towards succes with the nine-item, colour-ordering task.

Such results taken together provide a good deal of evidence to support the hypothesis that the spontaneous emergence of data reducing, economic, organisational structures occur, at least in the Cebus *apella* monkey, as a self-regulated response to a need to overcome the overwhelming cognitive strain otherwise effected by progressive increases in task difficulty.

Chapter 6

New areas of discussion and prospects for the future

Apart from the theoretical implications that these studies might ultimately have for comparative psychology, the experimental design and procedures which have been implemented and described in this thesis have given rise to the very first unambiguous demonstrations of the control of serial order productions of up to twelve-item sequences and have provided clear evidence for two-level hierarchical organisation in the non-human primate. Claims have also been made for the existence of data-reducing strategies on the part of the monkey subject, motivated by the need to be cognitively economic in the face of increasingly complex and difficult tasks given over a protracted period of some four years. Following the growth towards expertise of each individual monkey throughout a series of serial search tasks, indicants of a progressive adaptation to tasks of increasing levels of difficulty have continued to provoke the emergence of competences suggestive of a cognitive organisation previously unseen (or at least underdetermined) in traditional learning experiments in the laboratory.

However, from the outset of this open-ended 'voyage of discovery', there was never any guarantee that the monkey would succeed with such levels of task complexity as have been reported here for the first time. D'Amato and Columbo (1988) had only recorded a maximum five-item series previously (also with a *Cebus* monkey) and it was only with provision of the best conditions conceivable that we might have hoped for demonstrations of further success. Consistent daily performances have been produced now under supervised learning conditions in the laboratory with six adult monkeys housed in rich colony environments without any need for food deprivation schedules or other any less natural encouragement in order for them to engage in the experimental tasks. Indeed, the degree of perseveration has proven to be very high despite the increasing cost of failure, having progressed towards unprecedentedly long sequences in their later trials. Furthermore, as the tasks became steadily more difficult with these relatively long and increasingly complex sequences, the monkeys were continually working a lot harder for significantly less reward. This was so because irrespective of whether

the subject happened to be operating with a simple three-item string or a more complex nine-item fixed-order sequence, the trial success reward always remains to be a single peanut delivery via the feed-hopper. Indeed, the very success of the monkeys reported here are as much a result of the conditions under which the animals are housed and cared for as they are the outcome of the specific paradigms and procedures employed.

Nonetheless, the experimental paradigm as implemented here (McGonigle & Chalmers, 1996; McGonigle et. al., 1992, 1994) was also designed in such a way as to allow the informative interpretation of failure should the subject not be capable of such demonstrations. This was so arranged because the purpose was not so much in seeking to provide demonstrations of serial control of long sequences *per se*, but rather an attempt to provoke the development of behavioural adaptations which might open up the size of the working/decision space to be ideally large enough to allow its further experimental manipulation. Thus it would only be with such extended set sizes that a more affluent adaptive behavioural repertoire become available to the subject for further investigation with progressively more difficult taks.

The review of earlier comparative research provided in chapter one showed that there had continued to be a consistent failure to find any reliable dimension along which to meaningfully differentiate the 'higher' cognitive processes of different species (even with each other, let alone with humans). The writings of Darwin, Spencer and Romanes towards the end of the last century had set the scene for what was to become the study of comparative psychology, but lacking in any standard paradigms or procedures for reliable experimental observation, their work remained incomplete, largely anecdotal in nature, and not easily subject to replication. The methods of Thorndyke (1898), in an attempt to make more systematic the study of animal learning in the laboratory, were to mark a technical as well as theoretical contribution to the literature (his Puzzle-boxes providing a clearly reproducible environment for experimental manipulation with the same animal). However, possibly too complex to allow even their mechanical comprehension, lacking entirely from this situation was the means for the animal subject to possibly 'see' what the solution might be. It was only much later that Köhler (1925), working on "insight" in the chimpanzee, pointed out that for 'real'

learning to take place, some degree of forward-planning was required before the animal be able to perceive the instrumental value of an object as being a useful tool. Both his and the work of Yerkes (at the time trying to distinguish between 'insightful' versus trial-and-error learning curves) highlighted the issue of continuity within a session for the control of behaviour. In particular, it was in their experiments that it became clearly evident that in order to successfully come to the 'solution' to the stick-extension and box-stacking problems, the subject needed to perceive the spatial and ordinal relationships between items from a set of multiple alternatives. But although Thorndyke (1911) had proposed the importance of rational connectives for any development in mental life, no further empirical investigations were to follow up these ideas (even the later work of Harlow (1949) was to use completely arbitrary connectives). As a result, the search for some continuity in the mental processes of man and the other animals was not to be satisfied by means of the more introspective methods alone, and as a result lost ground to the then rising star of Behaviourism.

Following the perception of Pavlov's (1927) work as providing a more objective and scientific methodology, a more anti-mentalist era took hold and it was for Watson (1924) and later Skinner (1938) to suggest that much of what had at one time appeared to be 'intelligent' behaviour could actually be explained by an animal's history of reinforcement. Indeed, for Skinner, almost all behaviour could be predicted and controlled without reference to a mental life of any sort. However, unexpected failures of conditioning frequently occurred, and, as found by Breland and Breland (1961) for example, a pig taught to place a coin in a 'piggy bank' in return for a food reward, could later be found to merely 'play' with the stimulus (the coin) instead of placing it in the bank - even when food-deprived and motivated to work (see also Seligman (1970) concerning 'preparedness' responding). Although unwilling to support the view for a continuity between the cognitive processes of man and other species, Skinner was nonetheless right to highlight the importance of an individual's life-historical experiences and the effects of the dynamic interactions taking place between subjects and conditioning stimuli. Further, despite its shortcomings, with the invention of his 'Skinner Box', there was also the coming into being of yet another technological change which gave rise to a few welcome developments in methodological procedure. For example,

the technique of Skinner allowed the capture of very dense data sampling over extended periods with the same subject, often requiring the animal to have undergone certain prerequisite experimental conditions prior to a particular testing phase. Another important finding was that extinction was very poor for negative reinforcement because the animal did not have the opportunity to learn that the consequence was no longer to happen. However, this finding also highlighted another shortcoming of most of the traditional learning techniques, namely, that they failed to determine 'what was actually learned' in terms of the structure and content of learning over time, merely content to report the frequencies and rates of learning.

A further promising development was to come from the innovations of Harlow (1949) in which an animal subject was shown to derive quite profound effects upon its subsequent behaviour according to its own past experience with similar tasks. Using his newly developed WGTA apparatus (see figure 1.6) a subject was able to 'choose' among alternative stimuli in order to provide a response, and appeared to demonstrate 'learning to learn' by coming to 'know' the solution prior to the choice decision being made. He was also able to show broad species differences in performance with the learning sets employed (see also Passingham, 1981). However, as with so many of the other (now) traditional learning paradigms (using simple discrimination learning, matching to sample, etc..,) these tasks did not extend in level of difficulty and thus could not finely discriminate any definable cognate architectural differences in the putative structures or organisations of mental processes in different primate species (including man). Indeed in the case of the latter, the most widely held view (typically represented by MacPhail, 1982) was that any differences that were to be found (be they qualitatively or quantitatively measured) might be said to only reflect the possession of language.

Cognitive organisation does not require language

For the human cognition work, McGonigle and Chalmers had emphasised the continuing problems inherent in the use of experimental paradigms which heavily invest in complex language-like abstractions which could only be operated upon by the experimental subject who was already symbolically competent. The important consequence of these factors for

any comparative animal cognition study, is that the currency of explanation for many cognitive behaviours would start from a point beyond which most learning accounts finished, that was, the currency of symbol manipulation itself. If this was indeed the case, how then might we be expected to evaluate the existence of any 'higher cognitive' functions in the non-human animal ? The work of Piaget, although significant, had been of no help in this regard. His fame rested upon his developmental psychology, and in raising the status of the idea that development of any kind, not just psychological or behavioural, might be better understood in terms of within-organism, ontological evolutionary processes. Although Piaget's genetic epistemology centred around structuralist ideas of organisation, self-regulation, co-ordination and construction rather than the conventional evolutionary concepts of variation and selection, he was nonetheless unable to produce any convincing demonstrations of symbol-based cognitive development which did not presume a linguistic agent.

In contrast, the paradigms and procedures of McGonigle, employed throughout the seriation studies reported above, have involved the use of seriation tasks which did not require any verbal mediation. The actual string lengths and sequence composition conditions implemented were drawn from a larger number of possible decompositions available from the expansion of a simple [ABC] icon set, the background conditions of the longer sequences requiring antecedent successes so as to enable useful on-line monitoring and ongoing descriptive analysis as the monkey successively adapted to prospectively larger search spaces.

If any of the monkey subjects were to reach a level of expertise, say, with at least three items in each of at least two categories, once the hierarchical phase of the study was underway, it would quickly become evident whether the monkey would use class-based structure to help manage long sequences:

1. The first of these measures would be simply describable in terms of the length of sequences which these subjects might control when classification was indeed an option.
2. The second would be in observing the degree of spontaneity with which subjects search for items that were physically similar or resembled one another rather than continued to select items in a different category.

3. Thirdly, there would be an overall longitudinal pattern emerging from the life-history of each subject in the ongoing programme, over a number of years rather than months, indicating the degree to which an individual monkey might be coping with progressively more difficult problems in the face of reducing material rewards, with less cognitive cost.

Cognitive utility as measured by the disparity between objective measures of task difficulty and actual performance:

As the tasks can be objectively measured in terms of their increasing difficulty by computing their combinatorial expansion with each increase in sequence length, *ceteris paribus*, the strategic benefits which classification might appear to provide as a data reducing strategy are first indicated by accelerating acquisition rates as indicated in figure 6.1. (McGonigle and Chalmers, 1997). Indeed, outside the possibility of any chunking or other grouping of items together, the degree to which a monkeys' exploitation of a simple classificatory scheme might be to its advantage would be made transparent when their acquisition performance for varying sequence lengths be plotted against the combinatorial expansion for which the management of such sequence length increases otherwise entail.

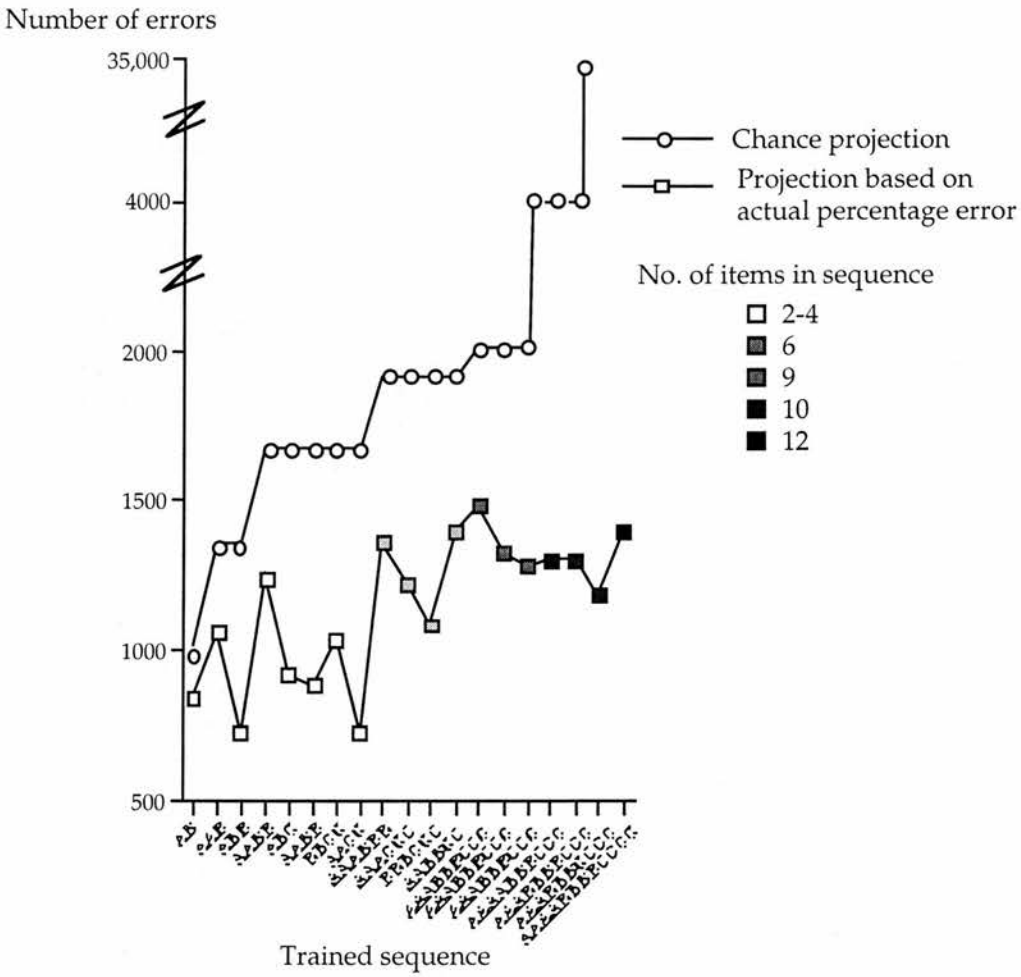


Figure 6.1. This graph is plotted in terms of the proportion of errors committed at each stage, where every data point represents a 75% success level in terms of correct productions that are non-reiterative, exhaustive and correctly sequenced. The graph shows two functions. One is based on the obtained (average) percentage error across different phases of the experiment but this score has been multiplied by 20 in order to compare the error functions with those we would expect on an assumption of chance - the second function shown - on a directly comparable scale. For this latter (projected) function, each item is treated as if it were independent from any other as would be in the case of an arbitrarily trained list, but adjusted to take account of the fact that the elements were sometimes identical (which makes chance projections less than they would be if every item were different - rather like a dice with e.g. three faces bearing the same number). Nevertheless on these adjusted 'combinatorial' projections, it is clear that the obtained functions are quite different in character from the chance ones. Rather than rise

exponentially with increasing sequence length, they show a clear plateau after the introduction of the first 9-item sequence. The learning functions do not therefore have the character they would show had learning simply been a matter of imposing constraint through brute force training, but show clear signs of divergence from this profile and thus subject-imposed constraint at successive stages of the task. These data represent the first record of a monkey's emergent classificatory schemes within a tutored situation.

(Taken with permission from McGonigle and Chalmers,1997, Oxford University Press)

The way ahead

In the light of the findings reported in the previous two chapters, it is now pertinent to address a set of questions, the answers to some of which have involved the development of new paradigms and procedures that require a level of expertise only now known to be within the capability of the monkey. A first issue concerns the *specific* utility of any organisation which has emerged over the course of training/learning. A second addresses the extent to which this organisation is the product of self-organisational processes. And thirdly, there is the comparative issue concerning which species can (also) exhibit the competences demonstrated by *Cebus apella*, and what other types of organisation one might expect to find in the animal world.

1. Cognitive utility: a direct measure ?

Having now quite clearly established that the monkey was capable of controlling the production of both linear (size) and categorical seriations, it was now conceivable that given these two different types of cognitive structure we could now evaluate their respective utility in a progressively larger search space. In a first attempt to achieve this aim it would be those lead subjects who had now completed the nine-item, two-level hierarchical series of experiments that would thereby qualify for exposure to a further set of supervised learning phases.

One implementation devised (McGonigle & Chalmers, 1997b) and presuming the expertise established by monkeys in the previous

experiments, confronts the subject with a seriation problem involving a six-item linear size-series. Starting with categorical seriation, a first transfer condition (see phase (i) below) would be one in which the monkey was confronted with six sizes and required to seriate these monotonically, where the first three size values are bounded by category [A], and the next three are bounded by category [B]. The two levels of size for each category would have to be initially tested in a transfer phase within which [$A_{s1}A_{s2}A_{s3}$] alternated with [$A_{s4}A_{s5}A_{s6}$], and [$B_{s1}B_{s2}B_{s3}$] alternated with [$B_{s4}B_{s5}B_{s6}$]. Following criterion runs for each of these transfer conditions, for which we would have been able to assess the extent to which the subject could base seriation on size relational codes, it would then be possible to test the following conditions for which the main contrast would be between categorical and linear seriation of six- (and if manageable) nine-item sets. As a first run, the subject would be presented with a three condition experiment (Phase (ii) below) consisting of two monotonic linear size seriations [A_{s1}-A_{s6}] and [B_{s1}-B_{s6}] (where each class is tested on its own with six differently sized items) and one categorical seriation [$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}$] (for which the six sizes were split into two halves, each represented by a different class). Should the management of these new seriations prove practical for the monkey, given that overall set-size plays such a crucial part in the emergence of organisational strategies, it would be most enlightening to see what kinds of sequencing characteristics were produced with such contrastive conditions with three-class, nine-item sets. This larger set size (Phase (iii)) would produce a contrastive case for three linear seriations [A_{s1}-A_{s9}], [B_{s1}-B_{s9}] and [C_{s1}-C_{s9}], with a single categorical seriation condition comprising three classes [$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}C_{s7}C_{s8}C_{s9}$].

Analogous to the 'comparison in the mind's eye' experiments of McGonigle and Chalmers (1984), such an experiment would allow a demonstration of the characteristics of search for both seriation and classification in the same subject, in the same experimental session, using conditions with equivalent numbers of items and level of difficulty. If there was to be any evidence for clear performative or organisational differences forthcoming between the linear sequences based on monotonic size series and categorical two-level seriation then quite *direct* measures of utility would have been provided for each, independently.

To illustrate that such direct measures of cognitive utility might be possible within the experimental programme overall, the following data is provided from a preliminary analysis of data derived from two monkeys which had more recently begun these later phases and were both in the six-item stage of this new experimental series.

Linear (size) and categorical seriation - acquisition phases for six-item sets.

Introduction

It was now possible or the first time to work with monkey subjects for whom it was certain that nine-item sequences were well within their capabilities, at least when given the option to classify the components into three distinct classes. Further, when each class was itself to have undergone the explicit ordering of its exemplars, these monkey were known to be capable of reporting consistent success with both size and colour variant compositionalities. However, this had so far only been investigated with category groupings for which a maximum single category membership of three items had been employed in any one condition. Whether such a performance might prove to be sustainable with an increased individual category membership remained in as yet uncharted territory, but served as a method by which the existing tasks could be extended, again both in terms of complexity and level of difficulty. Continuing our 'voyage of discovery' we returned to the same subjects with a viea to determining the extent to which the monkeys would continue to adapt their responses to the pressures of even greater non-arbitrary size coded seriations. The conditions of this new group of two-level hierarchical seriation experiments in a new Phase (i) were to be as shown in Table 6.1 below, and include the details of interspersed conditions and completion criteria for each phase. To prevent the subject taking too much new information on board at one time, the subsequent new sextuplet stimuli set exposures occurred over a staggered two phases; one categorical condition $[A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}]$, and two linear seriation conditions, $[A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}]$ and $[B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}]$ introduced in Phase (i)a_2, followed by sextuplet groups of $[A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}]$, $[B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}]$ and $[C_{s1}C_{s2}C_{s3}C_{s4}C_{s5}C_{s6}]$ to be introduced in Phase (i)b_2.

Table 6.1 Seriation Phases (i) [A B C]
Categorical and linear size seriation: 6 sizes/2 categories

Phase No.	Fixed order Conditions	Completion criterion
(i) a_1	$A_{s1}A_{s2}A_{s3}$ $B_{s1}B_{s2}B_{s3}$ $A_{s4}A_{s5}A_{s6}$ $B_{s4}B_{s5}B_{s6}$	15/20 15/20 15/20 15/20
(i) a_2*	$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}$ $A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}$ $B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}$	--- --- ---
(i) b_1	$B_{s1}B_{s2}B_{s3}$ $C_{s1}C_{s2}C_{s3}$ $B_{s4}B_{s5}B_{s6}$ $C_{s4}C_{s5}C_{s6}$	15/20 15/20 15/20 15/20
(i) b_2*	$B_{s1}B_{s2}B_{s3}C_{s4}C_{s5}C_{s6}$ $B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}$ $C_{s1}C_{s2}C_{s3}C_{s4}C_{s5}C_{s6}$	--- --- ---

*** 1 Categorical & 2 Linear Seriations.**

All of the monkeys participating in this new experimental series were to start with Phase (i) a_1 and thereafter, the monkeys were to follow the subsequent phase steps in the order as listed. This series of phases would require the monkey to attempt yet another two-level hierarchical series, but this time contrasted with a linear single category exemplar sequences in the same session, using an explicit size-ordering rule within each of the same three categories each to provide a six-item, six-size sequence. These six sizes might be spread either across two categories (e.g., $[A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}]$) or to be linearly arranged for one category alone (e.g., $[A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}]$. The six sizes to be used for each category would not include the same sized-icons as those employed in the original [ABC] sets for each monkey but retained those same category shape markers. The novel sizes covered the same overall size range as previously employed and were partitioned equally as measured by their geometrically central axis dimensions.

One possible solution to a merely exhaustive search of such arrays as presented in these phases might be to classify them on the basis of size alone, independent of shape, but in order to demonstrate the existence of a more controlled two-level hierarchical structure as outlined in figure 3.3 (and for comparison with this and the later extended linear size-order

conditions), the shape categories were to continue to serve as the 'superordinate' classes for a two-category condition under the present supervised learning conditions. The actual sequence compositions presented during these six-item size-ordering acquisition phases were drawn from a larger number of possible decomposition strings derivable from the [ABC] icon set, but were of a type familiar to each subject and conceivably simple in structure by comparison with other variants that may have been employed.

For every condition in this series, a unique linear, monotonic, increasing size-ordering was required to be reported by each subject for every icon set in the array, irrespective of the categorical compositionality of the icon set presented in each trial. Thus, for successful completion, every new array demanded an explicit and unique ordering of each and every individually identifiable icon upon the screen array. From trial to trial the same icons were to be presented in different configurations so as to avoid the production of stereotyped motor responses, whilst at the same time provoking continuous array interrogation on the part of the subject. The use of interspersed condition trials was also given partly for this reason, so preventing the subject's responses becoming too reflexive and thereby less 'cognitive' in production, whilst simultaneously allowing clearer interpretation of any error should there have been too great an increased task demand with the introduction of the new six-item, six-sized sets under supervised learning conditions.

Phase (i) a1

Rationale. Because the absolute sizes of the six new icons of each category would all be new to the animal, it was important to offer them initially as single triplet probe phases, presented as a start by four conditions: [A] and [B] category, 3-item sets comprising the first and second half of the six sizes as individual seriation tasks. Thus would any constant errors be better interpretable for the larger six-item linear size sequences if it be known that there were no problems concerning the icons being readily differentiable by their absolute size features.

Subjects. Two adult monkeys (Charlie and Mimi) took part in this phase.

Design and Procedure. The first probe trials for the new stimuli size sets were given as triplet groups of absolute sizes X_{S1}-X_{S3} and X_{S4}-X_{S6} for each of category shapes [A] and [B] so ensuring that each monkey was capable of discriminating the novel size differences to be used in the future seriation tasks. The four conditions were presented at a 1:1:1:1 ratio for all test conditions. The appearance of the screen array was such that all conditions contained one each of three differently sized monochrome contour shapes of categories [A] or [B] for explicit monotonic ordering within each class. Each triplet was to be interrogated in order of increasing size (e.g., touch X_{S1} before X_{S2} before X_{S3}) irrespective of the absolute sizes being displayed in the array. The phase was completed when all of the test conditions' criteria of fifteen correct out of the last (cumulative) twenty trials had been reached.

Results. Both subjects completed this four-condition phase successfully and were thus eligible for continuation with the two-level (six-size) hierarchical and linear phases of the experimental series. The mean number of trials, errors and percentage errors to criteria for the conditions were as shown in Tables 6.2-6.5.

Table 6.2 Nos. of trials, errors and % errors to criterion [$A_{s1}A_{s2}A_{s3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	51	4	7.8%
Mimi	81	58	71.6%

Table 6.3 Nos. of trials, errors and % errors to criterion [$B_{s1}B_{s2}B_{s3}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	49	5	10.2%
Mimi	25	10	40.0%

Table 6.4 Nos. of trials, errors and % errors to criterion [$A_{s4}A_{s5}A_{s6}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	49	10	20.4%
Mimi	15	5	25.0%

Table 6.5 Nos. of trials, errors and % errors to criterion [$B_{s4}B_{s5}B_{s6}$]

Subject	No. of Trials	No. of Errors	% Errors
Charlie	51	4	7.8%
Mimi	19	4	21.0%

Both subjects showed immediate successes with between 10%-80% correct trials out of their first twenty for any one condition, with order of completion varying between them. Analysis of the incorrect trials revealed no consistent pattern although typically more than 55% could be accounted for by a single one-step forwards-error touch [X_{s2}] for all conditions, despite there being four possible error paths executable. RT analysis revealed significantly decreasing time to be taken to touch the subsequent icons of every condition ($p<0.05$) for both subjects (see figure 6.2), typical of the profile seen for previous three-item phases.

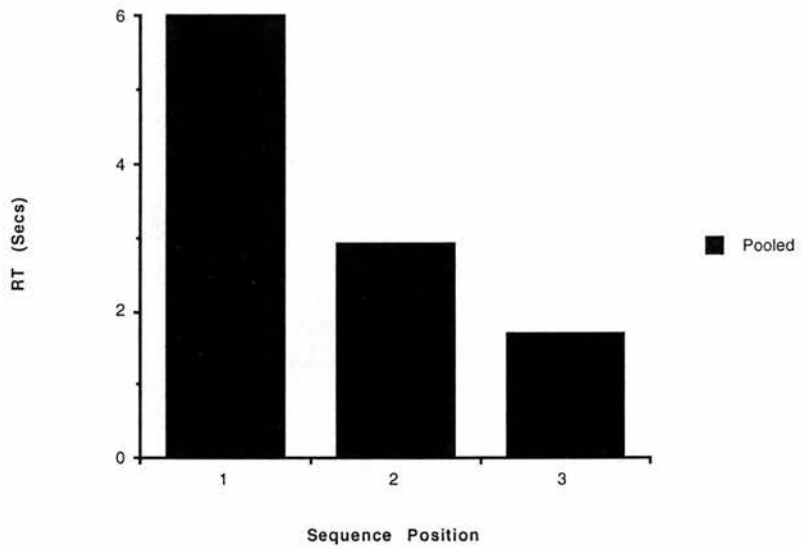


Figure 6.2 Mean RT profile for successful 3-item, single category fixed-order (size) sequencing by Cebus monkeys (N=2, pooled)

Discussion. It had now been established that each monkey would consistently report each of the category's two, three-sized item sequences independently of the absolute sizes of the icons to be explicitly ordered in each class. All conditions were completed with less time and effort (as measured by trials and errors to criteria) than that required to reach the same level of performance with any previous three-item set. Further, for the pooled data, most of the errors recorded were of one type, [X_{c2}],

suggestive of good generalisation of previous ordering patterns to these novel fixed-order sequences. Such preservations showed no evidence for sequential search ambiguity or other ordering difficulties. Confirming this trend, the RT findings of all three monkeys show that the time taken to search each subsequent icon of the set to be ordered, decreased throughout sequence execution, suggestive of the development of a route-planning strategy to be taking place. In support of this claim, it was noticeable that the initial reaction times (IRTs) were perhaps unusually large here compared to those for recent performances with size-ordering strings of much greater length.

Phase (i) a2

Introduction. This phase introduces the monkey to a set of interspersed categorical and linear (size) seriation sequences which would assess the subject's on-line management of serial order production with six independent items. The subject was required to exhaustively search and explicitly order by size, each and every item of any one of three six-item strings. Given that the inherent structure of a two category string was deeper than that of a linear, one-dimensional sequence, then the demands put upon working memory for sequences of sufficient comparable length would be different for the two cases. If correct, then one might expect that much more time and effort be required of the monkeys to reach an equivalent level of proficiency with a linear sequence than could be shown when other classification affordances were present. As well as determining acquisition rates, phrasing effects might also be affected in interesting ways that might inform us of the other differentiating involved in the monkey's control of serial order productions.

Subjects. Two adult monkeys (Charlie and Mimi) took part in this phase.

Design and Procedure. The first six-item, six-size sequences were offered in this phase with the new stimuli sets given as interspersed sextuplet groups of absolute sizes X_{S1}-X_{S6} for each of category shapes [A] and [B] in one of three ways. One categorical $[A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}]$ and two linear six-item conditions $[A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}]$, $[B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}]$ were presented, each composed of strings of increasing size. The three conditions were presented at a 2:1:1 ratio for the categorical to linear test

conditions, so ensuring equal numbers of trials for categorical and combined linear seriation exposures for comparison. The appearance of the screen array was such that all conditions contained one each of six differently sized monochrome contour shapes of categories [A] or [B] for explicit ordering within each class. Each sextuplet was to be interrogated in order of increasing size (e.g., touch X_{S1} before X_{S2} before X_{S3} ... X_{S4} .. etc,..) irrespective of the absolute sizes or classes being displayed in the array. No condition phase criteria were applied at first due to the very high level of task difficulty (even for the adult human subject), the performances being monitored and analysed on a daily basis, to be reviewed after 400 trial blocks over 8 sessions.

Results. Only one monkey (Charlie) has completed this phase to date, and his mean number of trials, errors and percentage errors to criteria for each condition were as shown in Tables 6.6.

Table 6.6 Nos. of trials, errors and % errors - Phase (i) a2

Condition	No. of Trials	No. of Errors	% Errors
$A_{S1}A_{S2}A_{S3}B_{S4}B_{S5}B_{S6}$	223	62	27.8%
$A_{S1}A_{S2}A_{S3}A_{S4}A_{S5}A_{S6}$	91	58	63.7%
$B_{S1}B_{S2}B_{S3}B_{S4}B_{S5}B_{S6}]$	88	14	14.0%

Both subjects showed immediate success with more than 75% correct trials among their first twenty. Analysis of the incorrect trials revealed a consistent finding across subjects, with unequal representation of the possible error types still in evidence, and, with different profiles being shown between the categorical and linear sequence error patterns. Accounting for between 70% and 92% of the total error trials completed for any one subject in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only five of them. Although Charlie's error profile continued to show many $[A_{S2}]$-1st touch errors (32%), the dominant error typologies otherwise reported for the categorical condition were $[A_{S1}A_{S3}]$ (43%), $[A_{S1}A_{S2}B_{S1}]$ (12%) and $[A_{S1}A_{S2}A_{S3}B_{S2}]$ (15%), each either a one-step forwards-category errors, or a one-step within-category touch error, with subjects recording very few reiterative touches (<1%). For the linear sequences, however, an additional error-type was commonly seen to occur in the fifth position $[A_{S1}A_{S2}A_{S3}B_{S1}B_{S3}]$ which could account for up to 30%

of the total error trials for those particular conditions. RT analysis revealed significant category boundary effects for both the subjects so far analysed, again showing differential profiles for categorical and linear sequencing as shown in figures 6.3-4 below. For Charlie, phrasing at the third position was significant for the categorical condition only ($p < 0.05$), coincident with the category boundary.

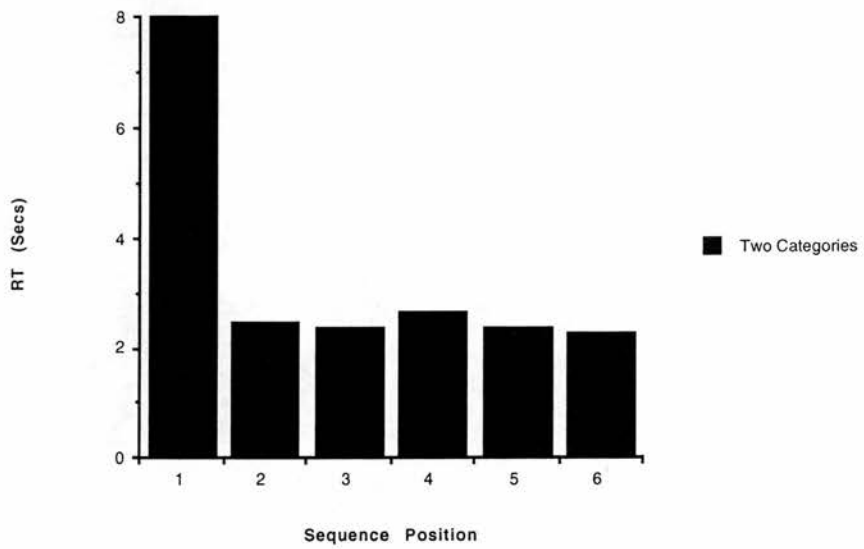


Figure 6.3 Mean RT profile for successful 6-item, two category fixed-order (size) sequencing by a Cebus monkey (Charlie)

Post hoc evaluations of the RT values for the linear sequence productions revealed no significant pairwise comparisons excepting that concerning the initial reaction time.

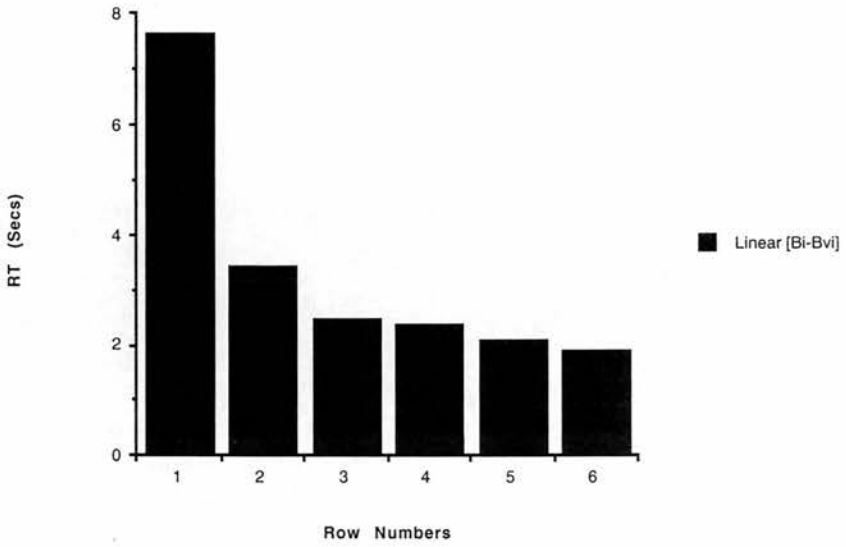


Figure 6.4 Mean RT profile for successful 6-item, single category fixed-order (size) sequencing by a Cebus monkey (Charlie)

The second subject to take part in this phase (Mimi, although still at an early stage with too few trials to allow significant data review) was also showing a tendency to differentially phrase between the two conditions in the same way, and at present, the effect is becoming more pronounced with each daily session.

Discussion. This phase has produced clear evidence that the monkey will successfully report a six-item, fixed size-ordered series, whether or not in the presence of classification options. Although of a high level of difficulty, immediate successes were shown by both subjects, showing more than 75% sequences in their first ten trials. Both error and RT profiles indicate unambiguous and clearly significant differences in serial productions of the categorical versus linear representations of the two conditions otherwise of equal length. The relatively higher success seen with the categorical condition is indicative of the same subject, in the very same session, taking advantage of the salient chunking possibilities of the two class condition (also reflected in the differential RT measures), a much more economic strategy than attempting a linear search, for which the increased number of items was now starting to place constraints upon working memory space.

Phase (i) b_1

To date, no subjects have yet undertaken this second set of probe phases for these novel sized three-item sequence evaluations. This phase was to provide a further check of the subjects ability to tolerate changes in absolute sizes of otherwise familiar icons, and repeats the methodology of Phase (i) a_1, but with categories [B] and [C] for the first time. Charlie had previously moved directly on to Phase (i) b_2 with the assumption that such tolerance would be met without any difficulty.

Phase (i) b_2

Introduction. This phase introduces the monkey to a further set of interspersed categorical and linear (size) seriation sequences which further assesses the subject's on-line management of serial order productions with six independent items. The subject is again required to exhaustively search and explicitly order according to a size rule, for each and every item of any one of three six-item strings. It was to be of interest to see whether the monkey would continue to show more time and effort being required to reach an equivalent level of proficiency with the linear sequences than would be shown when classification affordances were present. Further, it will provide us with a second opportunity to assess the reliability of the phrasing effects previously found to differ between the linear and categorical conditions during the monkey's successful control of both serial order productions in the same session.

Subjects. Only one adult monkey (Charlie) has taken part in this phase.

Design and Procedure. The second set of six-item, six-size sequences were offered as for Phase (i) a_2, with the new stimuli sets again given as interspersed sextuplet groups of absolute sizes X_{S1}-X_{S6} for each of category shapes [B] and [C] in the same one categorical and two linear six-item conditions, each composed of strings of increasing size. The three conditions were presented at a 2:1:1 ratio for the categorical to linear test conditions as before. The appearance of the screen array was such that all conditions contained one each of six differently sized monochrome contour shapes of categories [B] or [C] for explicit ordering within each class. Each sextuplet was to be interrogated in order of increasing size

(e.g., touch X_{S1} before X_{S2} before X_{S3} ... X_{S4} .. etc.,..) irrespective of the absolute sizes or classes being displayed in the array. No condition phase criteria were applied to the very high level of task difficulty (even for the adult human subject), the performances being monitored and analysed on a daily basis presented in 400 trial blocks over 8 sessions.

Results. Charlie has completed this phase, and his mean number of trials, errors and percentage errors to criteria for each condition were as shown in Table 6.7.

Table 6.7 Nos. of trials, errors and % errors - Phase (i) b_2

Condition	No. of Trials	No. of Errors	% Errors
$B_{S1}B_{S2}B_{S3}C_{S4}C_{S5}C_{S6}$	218	47	21.5%
$B_{S1}B_{S2}B_{S3}B_{S4}B_{S5}B_{S6}$	85	24	28.2%
$C_{S1}C_{S2}C_{S3}C_{S4}C_{S5}C_{S6}$	92	21	22.8%

The subject showed immediate successes with more than 85% correct trials amongst his first twenty for each condition. Analysis of the incorrect trials revealed a consistent finding to be found across conditions, with unequal representation of the possible error typologies still in evidence and different profiles to be seen between the categorical and linear sequence error patterns. Accounting for more than 92% of the total error trials completed for any one condition in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only three or four of them. Although Charlie's error profile continued to show many [B_{S2}]-1st touch errors (50%) for the categorical condition, they were recorded in negligible amounts for the linear searches (<8%). The dominant error typologies otherwise reported for the categorical condition were [$B_{S1}B_{S3}$] (15%), [$B_{S1}B_{S2}C_{S1}$] (23%) and [$B_{S1}B_{S2}B_{S3}C_{S2}$] (18%), each either a one-step forwards-category errors, or a one-step within-category touch error. For the linear sequences, however, an additional error-typology was commonly seen to occur, once more in the fifth position, [$B_{S1}B_{S2}B_{S3}C_{S1}C_{S3}$] which could account for up to 50% of the total error trials for those particular conditions. RT analysis again revealed a significant category boundary effect, again showing differential profiles for categorical and linear sequencing as shown in figures 6.5-6 below. Phrasing at the third position was significant ($p<0.005$), again providing a main rise-time effect only at the category

boundary. Other post-hoc evaluative comparisons revealed no significant RT differences for the linear conditions after the second position.

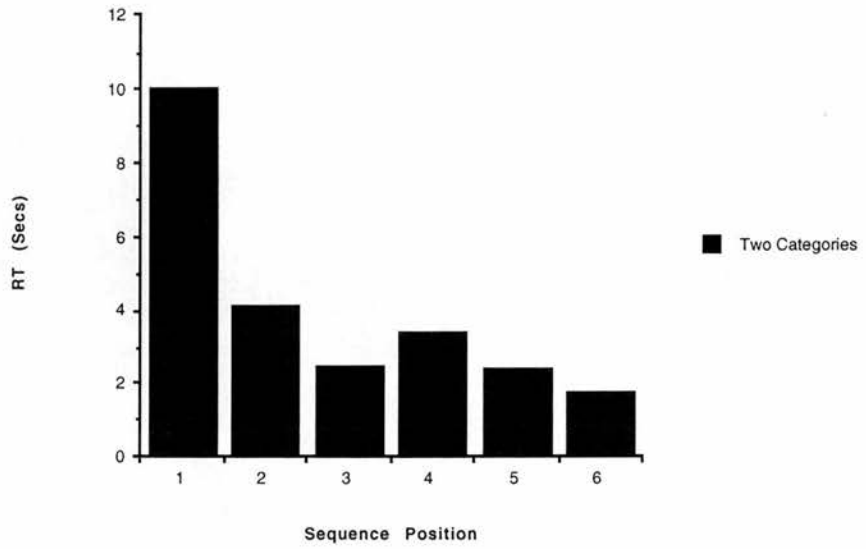


Figure 6.5 Mean RT profile for successful 6-item, two category fixed-order (size) sequencing by a Cebus monkey (Charlie)

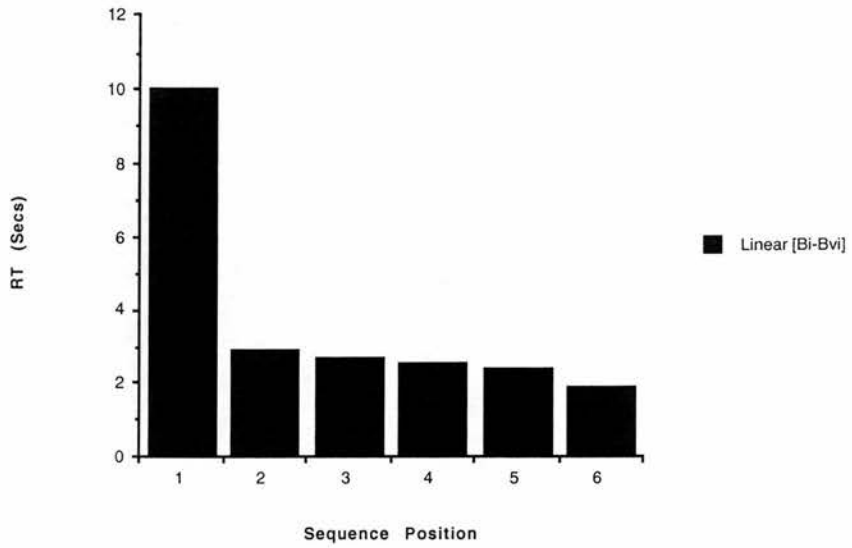


Figure 6.6 Mean RT profile for successful 6-item, single category fixed-order (size) sequencing by a Cebus monkey (Charlie)

Discussion. This phase has again produced clear evidence that the monkey would successfully report a six-item, fixed size-ordered series, whether or not in the presence of classification options. Both error and RT profiles indicate differences in serial productions for categorical versus linear representations of the two conditions otherwise of equal length and level

of difficulty. The relatively higher success seen with the categorical condition is again indicative of the subject taking advantage of the salient chunking possibilities of the two classes (as unambiguously and exclusively reflected in the RT measures), indexical of a much more economic strategy than when attempting linear searches, for which the increased number of items continues to place constraints upon memory space.

Linear (size) and categorical seriation - acquisition phases for nine-item sets.

Introduction

Now that the monkey had produced successful completions of six-item sequences requiring the explicit ordering of all their exemplars according to size (with and without the option to classify the components into distinct categories), the next step was to see whether such a performance might prove to be sustainable with yet a further increase in individual category memberships. Extending the task, again both in terms of complexity and level of difficulty, the conditions of this fourth group of two-level hierarchical seriation experiments were to comprise two final conditions at this stage, Phases (ii) and (iii), as shown in Table 6.8 below, detailing the interspersed conditions and completion criteria for each phase. To prevent the subject having to take too much new information on board at the one time, the subsequent new nontuplet stimuli set exposures occur over two phases; firstly some probe two-categorical/six-linear conditions were introduced in Phase (ii) to assure that the new absolute size values were both discriminable and could be reliably sequenced at all, followed by nine-item sets of three-category and single-category linear sequences to be introduced in Phase (iii). Completion criteria were not required for sextuplet groups of sizes [X_{S1}-X_{S6}] or [X_{S4} -X_{S9}] for any of the category shapes [A], [B] or [C] presented, but daily monitoring ensured that the monkey was capable of discriminating the novel absolute size differences for all eight conditions.

Table 6.8 Seriation Phase (ii) [ABC]
Categorical and linear size seriation: 6 of 9 sizes/3 categories

Phase No.	Fixed order Conditions	Completion criterion
(ii)*	$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}$	---
	$B_{s4}B_{s5}B_{s6}C_{s7}C_{s8}C_{s9}$	---
	$A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}$	---
	$B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}$	---
	$C_{s1}C_{s2}C_{s3}C_{s4}C_{s5}C_{s6}$	---
	$A_{s4}A_{s5}A_{s6}A_{s7}A_{s8}A_{s9}$	---
	$B_{s4}B_{s5}B_{s6}B_{s7}B_{s8}B_{s9}$	---
	$C_{s4}C_{s5}C_{s6}C_{s7}C_{s8}C_{s9}$	---

* = 2 Categorical and 6 Linear 6-item seriations

Categorical and linear size seriation: 9 sizes/3 categories

Phase No.	Fixed order Conditions	Completion criterion
(iii)*	$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}C_{s7}C_{s8}C_{s9}$	15/20
	$A_{s1}A_{s2}A_{s3}A_{s4}A_{s5}A_{s6}A_{s7}A_{s8}A_{s9}$	15/20
	$B_{s1}B_{s2}B_{s3}B_{s4}B_{s5}B_{s6}B_{s7}B_{s8}B_{s9}$	15/20
	$C_{s1}C_{s2}C_{s3}C_{s4}C_{s5}C_{s6}C_{s7}C_{s8}C_{s9}$	15/20

*= 1 Categorical & 3 Linear 9-item Seriations.

For each linear condition, a nine item set was to be presented for which each subsequent item of the set to be interrogated would be monotonically larger than its predecessor. The string items were to be exhaustively searched in order of increasing size for both categorical (three categories, three exemplar sizes of each) and linear (single category, nine exemplar sizes) seriation conditions.

Phase (ii)

Rationale. Because the absolute sizes of the nine icons of each category would all be new to the animal, it was important to offer them initially as single sextuplet probe phases, presented at first amongst eight conditions: [A], [B] and [C] category, 6-item sets comprising the first and second two-thirds of the nine sizes as individual seriation tasks. Thus would any constant errors be better interpretable for the larger nine-item linear size sequences if it be already nown that there were problems concerning the icons being readily differentiable by their new absolute size features. Contrastive categorical conditions were also given to assess the

significance of any continuing differential effects of the classificatory options.

Subjects. Only one adult monkey (Charlie) has to date taken part in this phase.

Design and Procedure. The first probe trials for the new stimuli size sets were given as sextuplet groups of absolute sizes X_{s1}-X_{s6} and X_{s4}-X_{s9} for each of category shapes [A], [B] and [C] so ensuring that each monkey be capable of discriminating the novel size differences to be used in the future seriation tasks. Two further conditions offered classification options with a requirement to maintain the strict six-item size ordering, right across the category boundaries [$A_{s1}A_{s2}A_{s3}B_{s4}B_{s5}B_{s6}$] and [$B_{s4}B_{s5}B_{s6}C_{s7}C_{s8}C_{s9}$]. The eight conditions were presented at a 1:1 ratio for all conditions, unless a change was initated as a result of consistent discrimination failure for any particular condition(s). The appearance of the screen array was such that all conditions contained one each of six differently sized monochrome contour shapes from the categories [A], [B] and [C] for explicit ordering within all classes. Each sextuplet was to be interrogated in order of increasing size (e.g., touch X_{s1} before X_{s2} before X_{s3}....X_{s6}) irrespective of the absolute sizes or shapes being displayed in the array. The phase was completed after daily monitoring had ensured that the monkey was capable of consistently discriminating and sequencing the novel absolute size differences for all eight conditions.

Results. To date, there has been no subject exit from this phase. The one monkey currently perfecting his performances has shown an equivalent standard as might have been required for him to have reached criteria for conditions independently, but he is not at present able to hold such a level of performance for all conditions simultaneously. The mean number of trials, errors and percentage errors for the first 550 trials were as shown in Table 6.8 below.

Table 6.8 Nos. of trials, errors and % errors - Phase 44

Condition	No. of Trials	No. of Errors	% Errors
$A_{S1}A_{S2}A_{S3}B_{S4}B_{S5}B_{S6}$	34	7	20.5%
$B_{S4}B_{S5}B_{S6}C_{S7}C_{S8}C_{S9}$	38	6	15.7%
$A_{S1}A_{S2}A_{S3}A_{S4}A_{S5}A_{S6}$	74	21	28.4%
$B_{S1}B_{S2}B_{S3}B_{S4}B_{S5}B_{S6}$	39	8	20.5%
$C_{S1}C_{S2}C_{S3}C_{S4}C_{S5}C_{S6}$	37	6	16.2%
$A_{S4}A_{S5}A_{S6}A_{S7}A_{S8}A_{S9}$	111	22	24.2%
$B_{S4}B_{S5}B_{S6}B_{S7}B_{S8}B_{S9}$	36	5	13.8%
$C_{S4}C_{S5}C_{S6}C_{S7}C_{S8}C_{S9}$	188	140	45.4%

The subject showed immediate successes with more than 50% correct trials amongst his first twenty for all except two conditions. The monkey appeared to be having difficulty not with sequencing *per se*, but did show some resistance with discrimination of the smaller-sized members of the [A] class, and the largest-sized exemplars of the [C] category (leading to ratio changes being activated). Analysis of the incorrect trials revealed consistent findings across conditions, with a nonuniform representation of the possible error typologies still in evidence. Differing profiles were also to be seen between the categorical and linear sequence error patterns. Accounting for more than 95% of the total error trials completed for any one condition in this phase, it was found that despite there being twenty-five possible error paths executable, most could be accounted for by only three or four of them. Although Charlie's error profile continued to show many [A_{S2}]-1st touch errors (60%) for the categorical conditions, they were recorded in fewer numbers for the linear searches (<10%). The dominant error typologies otherwise reported for the categorical condition were [$A_{S1}A_{S3}$] (10%), [$A_{S1}A_{S2}B_{S1}$] (10%) and [$A_{S1}A_{S2}A_{S3}B_{S2}$] (20%), each either a one-step forwards-category errors, or a one-step within-category touch error. For the linear sequences, however, an additional error-typology was commonly seen to occur, once more in the fifth position, [$B_{S1}B_{S2}B_{S3}B_{S4}B_{S6}$] which could account for up to 50% of the total error trials for those particular conditions. RT analysis revealed a significant rise time only at the category boundary for both categorical conditions ($p<0.005$), and a weak differential profile between categorical and linear sequencing (see figures 6.7-8). Post-hoc evaluations of the linear conditions showed weak phrasing effects at the third and fourth position were not significant for four out of the six linear conditions

($p<0.05$) providing pausing effects nonetheless, despite there being no category boundary. For three of the linear seriations, position five touches were significantly longer ($p<0.05$) than those adjacent elsewhere in the sequence.

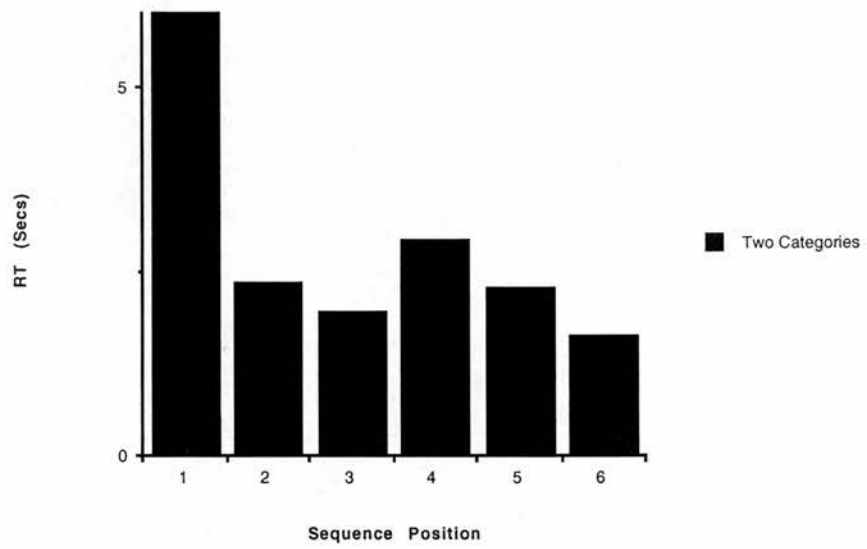


Figure 6.7 Mean RT profile for successful 6-item, single category fixed-order (size) sequencing by a Cebus monkey (Charlie)

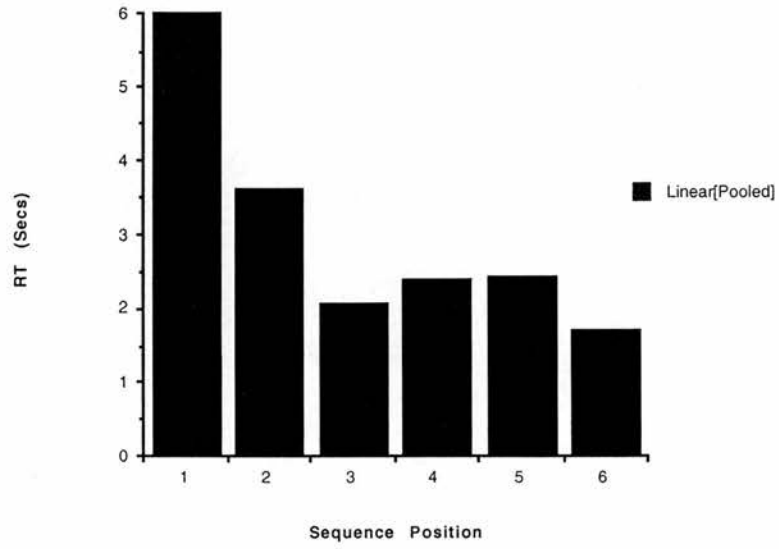


Figure 6.8 Mean RT profile for successful 6-item, single category fixed-order (size) sequencing by a Cebus monkey (Charlie)

Discussion. Although not yet to be completed by any monkey subject, this phase has already seen the demonstration of the sequencing of six-item sets using adjacent members of a nine-size monotonic order series. Early

successes were shown by the subject for most conditions, but despite there being some possible twenty-five error types for each of the conditions, error profiles reveal clear preferences for touching in accordance with the previous [AB] and [BC] adjacency sequencing, accounting for more than 80% of the error trials. These latter error-types were either one-step category forwards-errors, or one-step forwards-error seen within a category, each requiring its exemplars to be explicitly ordered. Category boundary effects were evident for both categorical and linear conditions with phrasing for the former showing unambiguous 'chunking' of the arrays into a three+three item sequence execution. We see here again, although still in its formative stages, that the monkey can demonstrate clear generalisation and transfer of their previously learned sequencing skills whilst provide good evidence for the existence and use of strategic planning with increasingly demanding and more complex sequences.

Further condition differences were to be found in the final stages of the experimental series to be so far reported, and were those seen to occur between the two kinds of six-item size-ordered searches for which classificatory options were not always present. Prior to this stage, any one category might have included a maximum of three exemplars for ordering (or possibly four in the case of free search within a class, see the earlier Phase 11). Given that there will be a limit beyond which any extension in breadth will itself start to impose memorial pressures within a category, it was possible to see whether such an effect could be initially investigated by offering the subject contrastive linear and categorical nine-item size-orderable sets. No subjects have as yet reached this stage, but a set of experimental phases involving six-sizes have been conducted and have been quite revealing. The pressures to produce a six-item linear sequence do indeed seem to pose greater strain upon the monkey subject (although they do succeed) than do they for the categorical condition as measured by the number of errors made. This is a reliable indicator for comparison in this instance, as each condition contains the same number of possible paths through their search space and the same number of error-types (and, *ceteris paribus*, are of equivalent levels of difficulty). What has been revealed (and only one monkey has so far moved on from this phase) is that not only does the reaction time profile differ for the categorical and linear conditions (only the former condition consistently showing both a category boundary and a significant phrasing effect, see figures 6.5-6) but

there is also a significant difference in the degree of errors produced. In the case of the conditions for which a linear size-ordering was required for all exemplars of a single category, there were almost twice as many errors committed than for those trials for which the classification of items into more than one group (by distinct class) was an option (see figure 6.9).

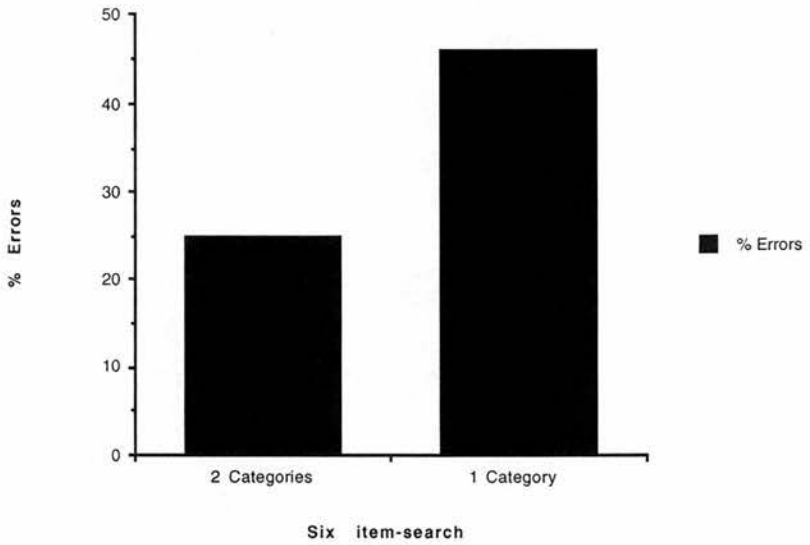


Figure 6.9 Percentage of errors for two six-item sets: (a) categorical- and (b) linear- size-ordered in a Cebus monkey (Charlie).

A further insight as to this seeming advantage of categorical affordances over the (albeit successful) management of linear sequencing is already becoming confirmed from casual analysis of one monkey's progress with six-item versions of a larger nine-size, nine-item search space (producing error findings similar to those above, but comparable RT profiles with increasing expertise). However, in the mean time, we shall have to wait for the results of a better test of the pressure to classify which is yet to come.

Conclusions: direct measures of cognitive utility

Following the successful implementation of this most recent set of procedures with the first monkey to exhibit sufficiently well adapated task experience, the relative utility of hierarchical and linear structures has appeared evident even with 6-item search spaces. Whilst at present represented by the current performances of only one monkey, it has been possible to simultaneously compare the effects of linear (one category, six

sizes) and hierarchical (two categories and six-sizes) ordering characteristics in the same subject within the same experiment. The monkey has demonstrated a great deal of success with both linear and hierarchical versions of a six-item set with the advantages of a hierarchical classification option making itself known by not only the relative numbers of correct trials, but also by the emergence of a self-organised principled production which is beyond the requirements of the task. Early indications are that chunking and phrasing effects remain strongly evident, providing further collateral for the claims made in the previous chapter for such phrasing to be indexical of a data-reducing strategy on the part of the monkey subject, motivated by the need to be cognitively economic in the face of an increasingly complex and difficult task. Following the growth towards expertise of each individual monkey throughout this series of serial search tasks, indicants of a progressive adaptation to tasks of increasing levels of difficulty have continued to provoke the emergence of competences suggestive of a cognitive organisation previously unseen in any comparative animal laboratory. If the monkey proves to be capable of ordering a nine-item linear (and/or) categorical sequence under supervised conditions (and there is no reason at present to suggest that it might not), then one implication of this is that the subject possesses a grasp of stimulus structure which is of considerable help in (usefully) constraining its search. If this proves to be the case, then one question which arises involves the extent to which the subject might be capable of self-organising search, for itself, on the basis of what it has learnt in the supervised phases already completed and described above. Such implications can also be tested within the programme on the basis of free search paradigms as devised by McGonigle and Chalmers, 1993. and now implemented with the monkey (McGonigle and Chalmers, 1997 a, b)

2. Self-organisation and free search as a direct measure of cognitive organisation.

This question of the degree of continuous adaptive change in the face of increasing levels of task difficulty is currently being explored by the use of a modified, free search and self-organisation paradigm developed from McGonigle and Chalmers' child seriation work in the wider context of the laboratory's comparative program of studies. In this recent development (McGonigle and Chalmers, 1997), the icons to be seriated were to be interrogated by the subjects along a free choice pathway of their own choosing, the only task requirement being for the production of a non-reiterative exhaustive search of the stimulus array. However, even though entirely novel search items might differ in their discriminable features, and be offered in the absence of any prior 'core' training designed to require a particular sequence, subjects would nonetheless retain the option of using a simple spatial strategy of the sort reported in an earlier spatial search task providing the first evidence of spontaneous self regulation in monkeys and children (De Lillo, 1994; McGonigle et. al., 1992).

In order to provide a task which did not require specific sequence training, whilst at the same time prevents the subject from using a simple spatial search heuristic, new procedures have been developed wherein the location of icons would be repositioned about the screen array as the subject was searching the set (McGonigle and Chalmers, 1997). What this task would then provide for the subject, would be a condition within which for each and every trial, following a correct (i.e., non-reiterative) touch to each of the subsequent icons of an emerging sequence, the absolute and relative positions of the array items discretely migrate about the screen array, thus removing the option of a positional cue being used as the basis for search on each individual trial. Such conditions require the subject to necessarily monitor the features of all the items in the array throughout the trial continuously, in the absence of any differential reinforcement for a particular success or failure to seriate all of the items of the array.

Preliminary results from pilot studies conducted with young children of around four years (by McGonigle and his colleagues in the laboratory) have shown this new task to be difficult beyond an array set of five items.

However, success has been recorded on the basis of icons discriminable by shape for up to and including an eight-item set (McGonigle and Chalmers, 1997). For those succeeding with the larger search tasks there is some evidence to suggest that they will spontaneously use, and maintain use of, a self-imposed order for all their sequences. On the other hand, the subjects who fail to successfully complete seriation with these larger arrays, are more idiosyncratic in their choice of pathway on a trial-to-trial basis, consistently changing their interrogations in no obviously principled way. As reported by McGonigle and Chalmers (1997), the extent to which the young human subjects move towards an 'approximation to a standard path has been calculated for all subjects, in which the observed number of order recall repetitions on two successive trials is subtracted from the expected'. Their preliminary results were summarised as shown in Figure 6.10 below and interpreted in a way convergent with data reported by Tulving (1962) who found that in a free recall task, it was for those subjects using a repeated order of report that the most success was achieved.

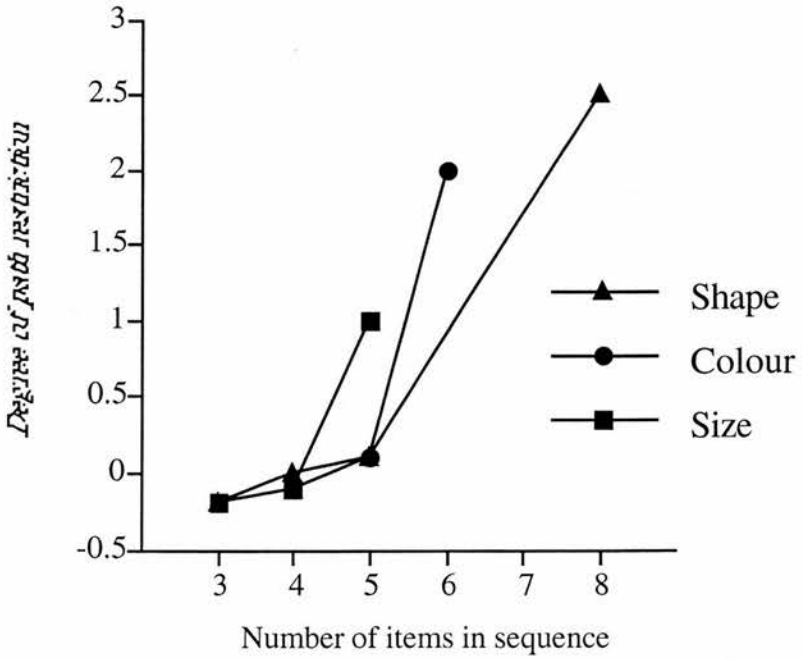


Figure 6.10 The relationship between path consistency and search success where the latter is measured by length of sequence produced without reiteration (reproduced with permission)

In the case of the size seriation task (the only condition for which there is a privileged path through the array) McGonigle and Chalmers (1997) report that despite the individual icon sizes being introduced into the

expanded sets on a basis of their random selection, it transpired that those subjects acquiring the most success on the longer sequences also repeatedly showed a preference for self-selective ordering of the set in a monotonic fashion. Similar analyses are currently being conducted for data acquired from naive monkeys in work now being extended by John Ravenscroft at the laboratory, and it will be interesting to see how comparable the results are, both when compared with children, and with the more experienced monkeys who have 'graduated' from the supervised experiments described in the main study of the thesis above.

Whether and which species might be capable of achieving these various levels and types of organisation may now be assessed in conditions of free search as summarised in figure 6.11.

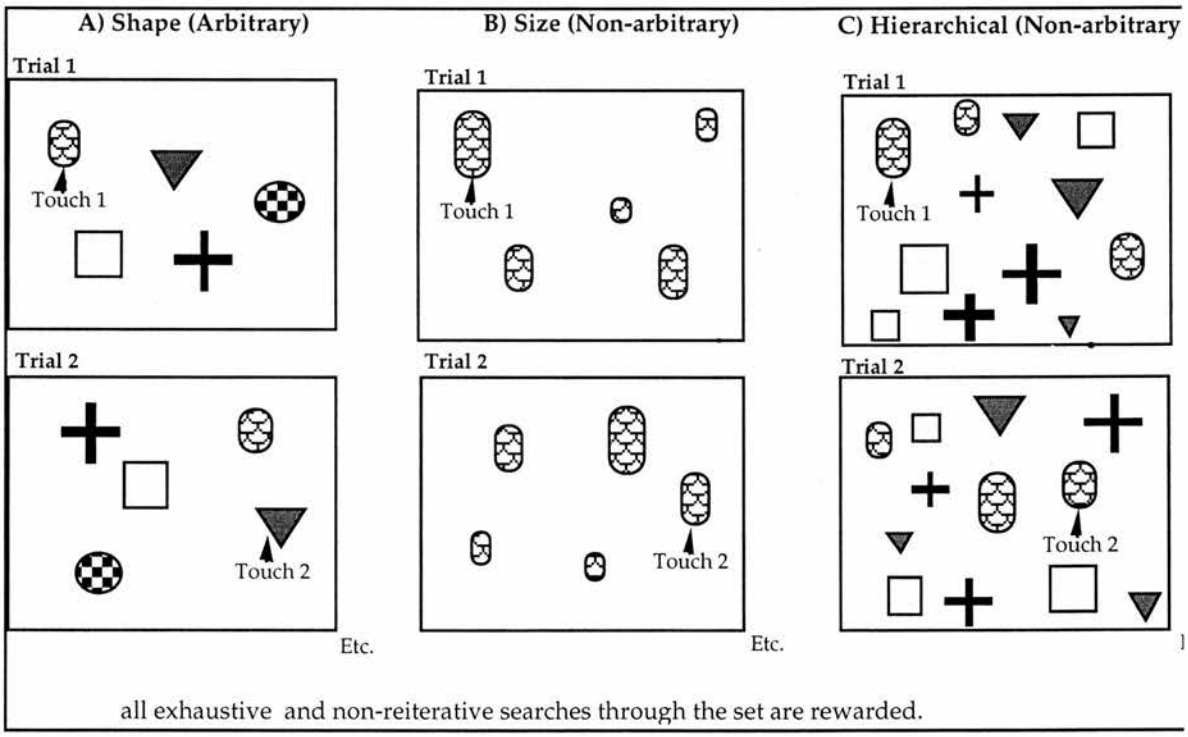


Figure 6.11 Basic free search paradigms of McGonigle and Chalmers (1997) (reprinted with permission)

The first of these tasks (see figure 6.11a) which presents icon sets discriminable by shape only, allows the subject the option of self-organising their search by 'devising a personal alphabet' of (arbitrary) items such that the search might be conducted on each trial in the same order. In the second (see figure 6.11b) in which essentially non-arbitrary

icon sets based on size relations are presented, the subject is able to use a principled and generalisable linear search strategy based on monotonicity cues from the array. Thirdly, in a spontaneous classification condition (see figure 6.11c) we can assess the use of a (general) categorical principle operating on the basis of grouping by e.g. shape then colour, as well as the degree of spontaneous hierarchical organisation; a further exploitation involving the possible ordering within a category.

In this way, these new paradigms and procedures would allow the fractionisation and independent evaluation of putatively different levels of organisation as discovered for both human and monkey subjects using a seriation task not mediated by language and which does not suppose the possession of any linguistic competence for its success.

3. Comparative implications for intelligent systems: an implementation with spatial search in the bird.

If, as McGonigle and Chalmers (1996) proposed, the emergence of economic search strategies were to be viewed as being indicative of epigenetic growth regulation in the face of increasing cognitive demands, then one might expect to be able to determine the degree to which a given agent might arbitrate between any number of possible solutions to a particular problem solely on the basis of economic behaviour. One paradigm specifically designed to address this question has already been implemented with both human and non-human subjects with encouraging results (McGonigle et. al., 1992). Using an exhaustive search task, the subject was required to (freely) interrogate each and every item displayed in an array. Up to nine physically identical items could be presented on a touch screen, with the specific icon configurations being changed on a trial-to-trial basis. The subject was free to search the items in any order they wished (including reiterative touching) in the absence of any explicit differential feedback for particular sequence executions. Given the entirely free choice of pathway through the items of the array, the degree to which the subject might demonstrate any economic management of the sequence could be measured by recording the total number of touches made prior to the completion of an exhaustive search of the set. For example, in the case of a nine-item set, the optimal number of touches would be nine, one touch to each icon, although rewards for both human

child (cartoon animation) and monkey subjects (peanuts) were provided for all exhaustive searches irrespective of the total numbers of touches made within the trial.

Using an incremental procedure, starting with a single icon and adding on one at a time until a total of nine items were reached according to the subject's continuing success, a first measure was to determine the number of icons manageable by each individual subject. Secondly, it was possible to see whether the subject might effect any changes in the actual paths chosen as each was repeatedly exposed to the same task over time. Such self-regulatory processes as might reveal themselves if the subject were to relax into any progressively more efficient search strategy, would be indicative of emerging cognitively adaptive responses unlikely to have resulted merely from a tightly coupled stimulus/response-type learning in an attempt to satisfy the requirements of the task.

As shown in figure 6.12, templates of both success and failure were produced in a first study with young children (McGonigle et. al., 1992) who showed clearly differentiable performances on the basis of age, with increasing set sizes. Not only were the older (four years old) children successfully searching larger arrays, they also showed a markedly more principled procedure for doing so with minimum reiteration.

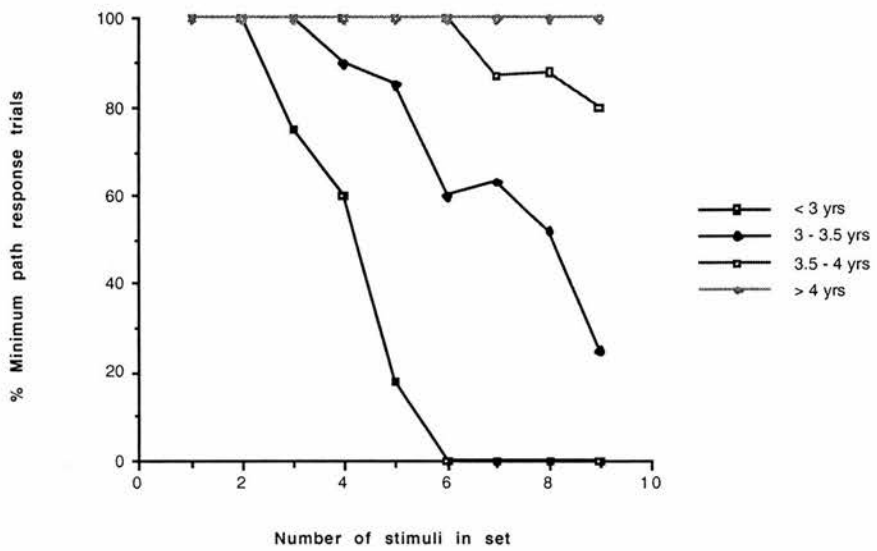


Figure 6.12 Search performance of young children in tasks of increasing numbers of identical items.

Based upon a vectorially-based touch sequence, these older children were consistently solving the nine-item sets by starting in a corner position, then moving either vertically up or down each adjacent column in a continuous snake-like manner as illustrated in figure 6.13 below. Even for the four-, five- and six-item sets, for which the configurational changes were more significant, there was a clear preference for choosing adjacent icons for sequencing following a left-to-right, top-to-bottom principle.

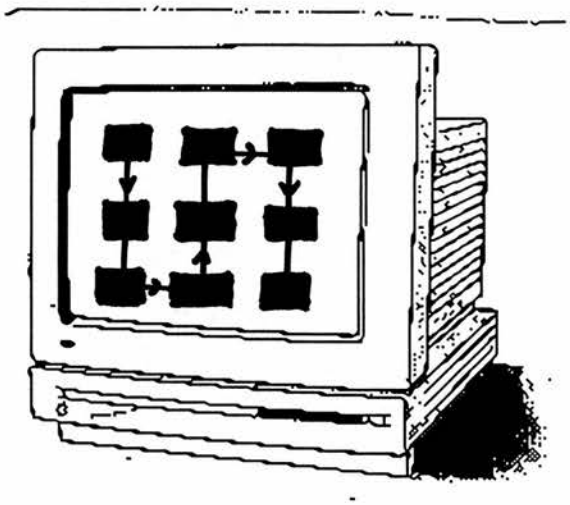


Figure 6.13 Example of self-imposed path restriction in the four year old child for an identical nine-item free search condition (line added to show sequential touch path)

Such a strategy appeared to become both more common and more consistent with increasing age, and was directly correlated with increasing levels of success with exhaustively searching the array. For the youngest subjects (from two years, eleven months) there were no such patterns of preferred path taking, and the sequence of touches actually made were typically random choice hits of icons in a variety of positions, often with large numbers of reiterative touching prior to completing the trial. This latter search behaviour not only showed these younger subjects to be uneconomic, they also became frustrated with their continuing failed attempts to successfully complete the exhaustive search condition of the larger sets.

Using the same paradigm and procedures with the adult monkeys in the same laboratory, DeLillo (1994; see also McGonigle et. al., 1992) report a similar economic path-finding development with increased exposure to

larger set sizes under free search conditions. Although not developing the consistent, regular minimal path strategy of the four year old children illustrated in the figure above, the monkeys nonetheless showed good evidence of self-regulatory behaviour based upon factors not provided by any immediate environmental arbitration. The evolving search strategies of the monkey proved very idiosyncratic, but all became progressively more efficient as the percentage of non-redundant touches increased with experience, and were similar to those of the young children in that the seriation of items in the array would often follow preferred starting positions and come to use adjacency principles. For the larger set sizes, the monkeys came to show patterns of touching which 'approximated to vectorial constraint' (McGonigle and Chalmers, 1996) although not to the same extent as that seen with the human subject.

Such ordering tasks have provided a useful window onto both the young human and monkey subject's abilities to improve their efficiency on the basis of their search 'experience' alone. If, therefore, the views of Macphail (1982) were to be correct, we might expect to see the same result with say, the pigeon, rat or cat, there being no obvious species differences in performance to be expected. With a view to addressing this question, another comparative study, also conducted within the same laboratory, assessed the abilities of avian subjects with the same task.

The original aim of this series of experiments was to assess the degree to which, if any, naive subjects would make economic movement through an item space of up to and including nine identical items arranged about a three-by-three grid. Given this task, any fully economic strategy would manifest itself as a minimal-path of touch response execution with neither repeated nor iterative touches per icon. If such a pattern of responding were to become more widely evident, as it has for both young children and *Cebus* monkeys, albeit vectorially or sequentially principled, this would provide the first evidence for auto-regulatory behaviour in the bird for an exhaustive-search multiple-item problem space.

Why the bird? The choice of the pigeon for inclusion in the comparative project was suggested by an existing literature of serial learning/serial search conducted with this species. In the last decades a series of experiments had been reported which described demonstrations of seriation skill performance in the non-verbal animal. Terrace (1986), using a three-element sequence of colour and achromat visual stimuli in a serial learning paradigm, had suggested that what the successful pigeon could have learned was the ordinal position of individual elements of a series. How ordinal position might be internally represented, however, remained a very important issue. Its transformation into a spatial array of information ordered associatively or by some relational means are likely postulates (Holyoak & Patterson,1981; McGonigle & Chalmers, 1986). In particular, following transitivity with the pigeon there was a reported maintenance of 'end marker' anchoring (Terrace, 1986). Intrinsic test difficulty was controlled for in transfer tests and it appeared that the use of a spatial code to define the ordinality of an element of a serial list looked likely as a principle of successful operation in this task. This notion suggested the development of spatial position experiments in which the elements of the series to be learned were not confined to any fixed position about the stimulus array. This development was put into practice with both four and five-item series lists showing successful pigeons to be utilising both 'end' anchors and 'chunking' of the element set for both whole-series and sub-set pairs sequential behaviour responses, (Terrace, 1991). However, what these findings did not tell us, was the degree to which a list might only be learned according to a series of conditional discriminators (having pressed [A] then prompts subsequent press of [B] which prompts subsequent press of [C] ..., and so forth) despite the evidence for chunking behaviour. Dallal and Meck's (1990) study is of interest here in that it had demonstrated that chunking could be obtained in serially organised behaviour in which an animal (in this case a rat in an Orton radial maze) was free to respond to the stimulus array in any order it chose. Terrace's pigeon work with three, four and five-item lists confirmed the reliability of lists composed of heterogeneous and segregated items as vehicles for establishing chunks, but these findings were derived from protocols which ultimately require fixed-item series orders, many of which may

have been completed using end markers as their default rules (Terrace and Chen, 1991a).

As with analyses of chunking in human memory (Miller, 1956), studies of chunking in animals had often assumed that its function was to reduce memory load on the working memory and thereby make it easier for the animal to produce an arbitrary list. Having found chunking to be evident for four-item lists (Terrace & Chen, 1991b), one might suggest that for the pigeon, multiple-item set for serial learning would be well beyond the working memory load required for success at a seriation task of even moderate complexity. With these considerations in mind therefore, there was no *a priori* reason to suggest that a pigeon might be capable of making successful movement through a homogeneous item space of up to nine items at all, let alone that it show an economy of movement as optimised by exhaustive search task success. Having presented the shaped and equipment pre-trained pigeon with this task, any success measured was to be described in terms of spontaneous classification of the item space array by the pigeon, a claim previously discountable on any number of grounds, some of which have been discussed above.

Design and procedure. All of the eight birds (*Columba livia*) participating were initially naive concerning the aims of the study, none of them having seen the stimulus array nor apparatus previously. By the end of the experiment to be briefly reported here, up to seven phases had been completed, the involvement of any one subject at each phase being determined by their progressive performance assessment as for the monkey.

Stimuli. The stimuli employed were identical to those used with monkeys and children with only those modifications deemed to be necessary for reasons of ergonomic and physiological comfort. As stimulus material, up to four green coloured 'squares' (35 x 40 mm) were randomly positioned and presented within a symmetrical three by three matrix (170 x 190 mm) in the centre of an interactive video touch screen. Pixel densities were constant for each target generated by a graphic system, the software of which allowed automatic recording of touch screen response data in real time. The data obtained by this method included all touch latencies and sequence of touch information including those of repeated and reiterated

target responses. Subjects were situated in a purpose-modified Skinner box with one side opening onto the interactive touch-screen at a bench height adjustable to the bird's required comfort when operating the system. Behaviours (on and off screen) exhibited by every subject during each experimental trial was recorded on video tape for off-line analysis.

Pre-training. At no point were the birds task-informed other than that they were to initially approach the screen with a forward-facing head and beak. Reinforcement (free seed) was offered for successive approximations towards icon-pecking postures once screen-positioned. Once the bird was pre-trained to the apparatus, differential feedback was never provided for touch sequence, latency or any other performance parameter, prior to exhaustive touching each of the item set shown on the screen during each array presentation.

Testing procedure. In each phase of the experiment the pigeon was required to interrogate a touch-screen array and to sequentially interrogate every icon presented as for the *apella* monkey (see also DeLillo (1994) for free search comparison). All subjects once comfortable in the test environment were presented with a single 'square' and encouraged to make a touch response towards it with an outstretched beak. At no time throughout the experimental series was the subject given any explicit tuition or had other experimenter interaction (although present). Having touched the target square, a tone would sound and the target would disappear from view for 200 mS. For this and all screen displays in this experimental series, once each of the targets had been touched, the array was removed from the screen. If, and only if, the touch sequence showed neither repetition nor reiteration of any of the target items in the array, free food was made available for 5 seconds from a food-hopper situated to the left of the touch-screen. If, however, a subject showed either repeating or reiteration of target touch responses for the current screen array, the screen became blanked white for 30 seconds, the feed-hopper remaining inactive. Subjects were taken through incremental item number arrays of the series, from 1-4 items per screen array, dependant upon their degree of success and task-motivation as determined by the experimenter. Each pigeon qualified for continuation by either (i) a criterion of 70% successful seriation in two consecutive sessions or (ii) a plateau (or decline) of performance over 120 trials. At no time was

differential feedback offered to any subject during the exhaustive search series of experiments. Each bird was run individually in a pre-determined and fixed order at the same hour of the day, five days per week. The experiment was automated for both stimulus presentation and feed access components, taking place in a closed room with subdued lighting and a background of white noise. A daily session would contain a maximum of 40 trials.

Results. The overview of results given here reports the progress of 8 birds through an experimental series of incremental 1-4 homogeneous item arrays, and 1-4 item, two category arrays.

Of the eight adult pigeons participating in the experiment, all successfully completed a one item set within two weeks of equipment adaptation and screen-pecking behaviour acquisition [mean = 7.25 sessions, range = 5-13]. For the two-item series, six out of the eight birds reached criterion for the phase, one reached a plateau of performance at less than 50% minimal paths per session, another was removed from the experiment due to its developing disinterest after some 500 trials. In the three-item phase, seven birds completed at least 1000 trials each. Only one achieved the 70% minimal path success criteria over more than one session, the other six showing a plateau followed by a declining performance after ten sessions. Only three birds entered the four-item set condition, none of which ever achieved a minimal path performance greater than the 25% level for any one session. Indeed, success by this criteria dropped for all three birds after their first 300 trials and further still after some 1000 trials.

Condition	No. of birds (Max = 8)
A	8
AA	6
AAA	1 (Max = 7)
AAAA	0 (Max = 3)
AB	3 (Max = 3)
AAB	2
ABB	2
AABB	0

Figure 6.2 Numbers of birds reaching criterion levels of performance for different 1-4 item sequences.

For the fixed-order classification condition all three pigeons succeeded with attaining the 70% criteria for a two-item sequence, but found the

addition of a third to require considerably greater effort to be required. For a four-item sequence (two categories, with two exemplars each) no subject was able to maintain a maximum 35% correct trials over two sessions.

Discussion. Although by these measures the pigeon performances do not offer support for spontaneous classificatory behaviour (and all RT measures for individual sequence position showed too wide a variance to be reliably used), it is not true to say that the birds did not start to self-regulate themselves towards the task in any principled ways. For example, the high degree of persistence shown following failure alerts one to the continued motivation to task which these subjects maintained. Initial reaction times to the first touch decreased with time and the decrease in the number of repeated pecks to previously touched icons was quite dramatic after the two-item set phase. One of the more striking illustrations of the bird's change in behavioural orientation following task success was seen in the decreasing frequency of false positive anticipatory movements towards the feeder prior to task completion. This occurred not only within a phase (not so surprisingly) but between phases of increasing sequence length, indicative of the bird's changing both search and response strategy. Further developments were evident from the use of spatial and adjacency cues in the execution of the bird's response pattern under free search conditions. Although idiosyncratic in nature, seven out of the eight pigeons showed a tendency to interrogate the array in a left-to-right and top-to-bottom manner. Caution should be noted here, however, as the use of a left-to-right movement across the screen could have merely reflected a proximity effect due to the feeder position.

So, did repetition for the task lead to economy of search and/or other strategic, adaptive change for the bird ? As measured by percentage minimal paths, after some 2000 trials for each bird the answer was negative for a homogeneous four-item set using the free search paradigm, the pigeon seemingly unable to remember which icons it had already visited when more than three were present. For the categorical condition, in contrast to the findings of the *apella* monkey, there appeared to be no advantage taken by the bird of the classificatory affordances of the array with increasing set size. Indeed the supervised and non-supervised

versions of the task told the same story for the avian subject, with the bird losing track of their place in the sequence when attempting to search more than three to four items.

Although perhaps confounded by the birds tendency to move towards the feeder after icon touching at first, path restriction in the spatial study resonates the results of the study with young children and is analogous with the path restricting, data reducing procedures (or lack of them) as seen in the visual classification and seriation experiments with children and monkeys. The free search paradigm as implemented here has proven to be very sensitive to the discovery of both quantitative and qualitative differences both within and between species, providing measures of non-trivial aspects of cognitive growth and development.

Generality and stance:

In summary, the experiments reviewed above were based on a radical paradigm shift from conventional learning experiments which have consistently failed to capture the more dynamic aspects of a comparative program in their attempts to discover the evolution of intelligence as seen in the sophisticated human adult subject. It has been shown that, under conditions for which sufficient choice of response may be made from an otherwise large set of alternative responses in order to satisfy task demands, given repeated exposure, a subject able to monitor its own behaviour is capable of making progressively economic decisions on the basis of the amount of effort required to achieve success. Although the consequences of McGonigle and Chalmers' use of non-verbally mediated tasks have proven a resounding success in use with both young children and monkey subjects, it is not merely the demonstration of self-regulated, controlled, structured performance criteria that they seek. Rather, their position is grounded in a more evolutionary biological program of research within which they seek to address the question as to 'what kinds of action and activity are really important in enabling complex systems to develop epigenetically through a space of possibilities ?' (McGonigle & Chalmers, 1997a p.21) Indeed, it has always been the contention of McGonigle (pers. comm.) that one dimension along which different species might meaningfully be compared was the extent to which a particular organism was 'able to arbitrate and self select the least resource

demanding procedures to achieve the maximum task adaptation with the least effort' (see also McGonigle & Chalmers, 1997b p.1). By such criteria, one perhaps warrants the claims made above for the use of classification in both the human child and the monkey as exploiting 'meaningful' self-imposed constraint for the purposes of organising and negotiating a path through a complex array. This is, however, not a given universal for all species, and a gradeable response has indeed been made evident using free search procedures.

The experimental paradigm as implemented here (McGonigle, 1990; McGonigle et. al., 1992, 1994) was designed in such a way as to allow the informative interpretation of failure should the subject not be capable of such demonstrations. This was so arranged because the purpose was not so much in seeking to provide demonstrations of serial control of long sequences *per se*, but rather an attempt was being made to provoke the development of behavioural adaptations which might open up the size of the working/decision space to be ideally large enough to allow its further experimental manipulation. Thus in the case of seriation, it would only be with such extended set sizes, putting continuous pressure on the subject to produce successive adaptive responses in the face of increasing task difficulty, that a more affluent behavioural repertoire might become available for further investigation. This demand for a more dynamic interplay between the subject (of whatever species) and the changing task environment was a feature that had been left wanting in the more traditional learning experiments, and it should therefore come as no surprise that the 'levelling effect' so often seen from the results of short-term snapshot studies came to find so few species differences. If one is seriously attempting to assess the growth and development of cognitive competences, then it will be necessary to conduct longitudinal studies which provide sufficient opportunity for individual subjects to self-regulate their own behaviour over time, using indicants based upon successful performances in response to continuously evolving challenges from the (task) environment. If such is not the case, then there is essentially no metric being provided along which a subject's growth trajectory for a given competence may be measured.

The paradigms and procedures of McGonigle and Chalmers (1997a) lay bare a program for the discovery of situated, adaptive intelligence as may

emerge from an ongoing series of studies (of which the work of this thesis comprises only a recent episode) designed to fractionate complex cognitive skills into their more basic competences without the use of language-mediational of other linguistically-based tasks. Using a search based metric of task difficulty, and evaluating cognitive organisation as a way in which an agent constrains the search space and reduces the difficulty of the problem as a consequence, it has been possible for the first time to directly compare the nature of the dynamic interaction between status of agent, its life history, and the opportunities it has been given to adaptively compensate for progressive increases in task difficulty. In this way, a multidimensional space has been created which enables the assessment of species comparative and developmental cognitive assays using a common currency of both task and measurement.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought.* New Jersey: Lawrence Erlbaum

Banks, W. P. (1997). Encoding and processing of symbolic information in comparative judgements. In G. Bower (Ed) *The Psychology of Learning and Motivation* Vol. II. New York: Academic Press.

Boakes, R.A.(1984). *From Darwin to Behaviourism: psychology and the minds of animals.* Cambridge: Cambridge University Press.

Breland, K. & Breland, M. (1961). The Misbehaviour of Organisms. *American Psychologist,* **16**: 681-684

Bruner, J. S., Olver, R. R. & Greenfield, P. M. (1966). *Studies in Cognitive Growth.* New York: Wiley.

Bryant, P. E. (1974). *Perception and understanding in young children: an experimental approach.* Methuen, London.

Bryant, P. E. & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature,* **232**: 456-8.

Byrne, R.W. & Whiten, A. (1988). *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans.* Oxford: Clarendon Press.

Byrne, R.W. (1993) Do larger brains mean greater intelligence ? *The Brain and Behavioural Sciences*, **16**: 696-697.

Byrne, R.W. & Whiten, A. (1997). Machiavellian intelligence. In A. Whiten & R.W. Byrne (Eds) *Machiavellian intelligence II: Extentions and evaluations.* Cambridge University Press.

Chalmers, M. & McGonigle, B. O. (1984). Are children any more logical than monkeys on the five term series problem ? *Journal of Experimental Child Psychology,* **37**: 355-377

Chalmers, M. & McGonigle, B. O. (1996). Capturing dynamic self-regulation in the laboratory. In L. Smith (Ed.) *Piaget, Vygotsky and Beyond*. London: Routledge

Changeaux (1990). *Neuronal Man.* London: Fontana Books.

Cheney, D.L. & Seyfarth, R.M. (1990) *How Monkeys See the World: Inside the Mind of Another Species.* University of Chicago Press.

Chomsky, N. (1959). "A review of B.F. Skinner's *Verbal Behaviour"* In J. A. Fodor & J. J. Katz (Eds), *The Stucture of Language: Readings in the Philosophy of Language,* 547-578. Englewood Cliffs, NJ: Prentice Hall.

Chomsky, N. (1972). *Language and Mind.* New York: Harcourt Brace Janovich.

Chomsky, N. (1980). *Rules and Representations.* Oxford: Blackwell.

Cole, R.W. (1907). Concerning the Intelligence of Racoons. *Journal of Comparative Neurology and Psychology,* **17**: 211-61.

Craik, K.J.W. (1943) *The Nature of Explanation.* Cambridge University Press. Cambridge.

D'Amato, M. R. & Colombo, M. (1989). Serial learning with wildcard items by monkeys (*Cebus apella*): Implications for knowledge for ordinal position. *Journal of Comparative Psychology,* **15**: 252-261.

Dallal, N. L. & Meck, W. H. (1990). Heirarchical structures: Chunking by food type facilitates spatial memory. *Journal of Experimental Psychology*, **16**: 69-84.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection.* London: Murray.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex.* London: Murray.

Dawkins, R. (1989). *The Selfish Gene.* New York: Oxford University Press.

De Lillo, C. (1994). The Logic of Memory Search in Non-human Primates (*Cebus apella*). Unpublished Ph. D. Thesis, Edinburgh University.

DeSoto, C.B., London, M. & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology,* **2**: 513-521.

De Waal, F. (1989) *Peacemaking among Primates.* Cambridge, Mass. Harvard University Press.

Dickinson, A. (1980). *Contemporary Animal Learning Theory.* Cambridge: Cambridge University Press.

Dickinson, A. (1985). Actions and Habits: the development of behavioural autonomy. In L. Weiskrantz (Ed.) *Animal Intelligence, 67-78.* Oxford: Clarendon Press.

Dunbar, R. (1997). *Grooming, Gossip and the Evolution of Language.* London: Faber

Edelman, G. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind.* New York: Basic Books.

Etienne, A. S. (1973). Developmental stages and cognitive structures as determinants of what is learned. In R. A. Hinde and J. Stevenson-Hinde (Eds) *Constraints on Learning.* London: Academic Press.

Flavell, J. H. (1963). *The Developmental Psychology of Piaget.* New Jersey: Princeton.

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition,* **28**: 3-71

Gardener, R. A. & Gardener, B. T. (1969). 'Teaching sign language to a chimpanzee' *Science,* **165**: 664-672

Garnham, A. (1985). *Psycholinguistics.* Oxford: Blackwell.

Gould, S.J. (1977). *Ever Since Darwin.* New York: W. W. Norton.

Greenfield, S. (1997). *The Human Brain: a guided tour.* London: Wiedenfeld & Nicolson

Gregory, R. (1984). *Mind in Science.* London: Wiedenfeld and Nicolson.

Griffin, D.R. (1978) Prospects for a cognitive ethology. *The Behavioural and Brain Sciences,* **4**: 527-538.

Griffin, D.R. (1982, [1976]) *The Question of Animal Awareness*:: *Evolutionary Continuity of Mental Experience* (Revised Edition) New York: Rockefeller University Press.

Griffin, D.R. (1984) *Animal Thinking.* Harvard University Press.

Griffin, D.R. (1992) *Animal Minds.* Chicago University Press.

Harlow, H. (1949). The formation of Learning Sets. *Psychological Review,* **56**: 51-65

Hays, N. (1994). *Principles of comparative Psychology.* Sussex: Lawrence Erlbaum.

Hayes, K.J. & Hayes, C. (1951) The intellectual development of a home raised chimpanzee' *Proceedings of the American Philosophical Society:* **95**: 105-109.

Hendriks-Jansen, H. (1995). *Catching Ourselves in the Act: situated activity. interactive emergence, evolution and human thought.* MIT Press. Cambridge, Mass.

Herrnstein, R.J. (1979) Acquisition, generalisation and discrimination: reversal of a natural concept. *Journal of Experimental Psychology: Animal Behaviour Processes*, **5**: 116-125.

Herrnstein, R.J. & de Villiers, P.A. (1980) Fish as a natural category for people and pigeons. In G.H. Bower (Ed.) *The Psychology of Learning and Motivation.* Vol. 14: 60-97. New York: Academic Press.

Hobhouse, L.T. (1901). *Mind in Evolution.* London: Macmillan.

Hodos, W. (1982). Some Perspectives on the evolution of Intelligence and the brain. In D. R. Griffin (Ed.) *Animal Mind-Human Mind.* London: Springer-Verlag.

Holyoak, K. J., & Patterson, K. K. (1981). A positional discriminability model of linear-order judgements. *Journal of Experimental Psychology: Human Perception and Performance*, **7**: 1283-1302.

Hull, C. L. (1945). The Place of Innate Individual and Species Differences in a Natural Science Theory of Behaviour. *Psychological Review,* **52**: 55-60.

Hunter, W.S. (1914). The Delayed Reaction in Animals and Children. Reprinted in W. Dennis, (Ed.) *Readings in The History of Psychology*, 1948. New York: Appleton-Century-Croft (Orig: *Behaviour Monographs,* **2**: No.6. Pubs: Williams & Wilkins).

Inhelder B. & Piaget, J. (1964). *The early growth of Logic in the Child.* London: Routledge and Kegan-Paul.

Jennings, H.S. (1906). *The Behaviour of the Lower Organisms.* New York: Macmillan.

Jerison, H. J. (1969). Brain Evolution and Dinosaur Brains. *The Amateur Naturalist,* **103**: 575-88.

Jerison, H. J. (1973). *Evolution of the Brain and Intelligence.* New York: Academic Press.

Karmiloff-Smith, A. (1991). Beyond Modularity. In S. Carey & R. Gelman (Eds) *The epigenesis of mind.* Hillsdale: Lawrence Erlbaum.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Change.* London. MIT Press.

Karmiloff-Smith, A. (1993). Self-organisation and Cognitive Change. In M. H. Johnson (Ed.), *Brain Development and Cognition: A Reader.* Oxford: Blackwell.

Kellogg, W.N. (1968) Communication and Language in the Home-raised Chimpanzee. *Science* **162**: 423-427.

Kellogg, W. N. & Kellogg, N. A. (1933). *The Ape and the Child.* New York: McGraw-Hill.

Klüver, H. (1933). *Behaviour Mechanisms in Monkey and Man.* Chicago: University of Chicago Press.

Köhler, W. (1925). *The Mentality of Apes.* (Translated: E. Winter of 2nd Germ. edn. 1917) New York: Harcourt, Brace & World.

Krechevsky, I. (1932). "Hypotheses" in Rats. *Psychological Review,* **39**: 516-532.

Krechevsky, I. (1938). A study of the continuity of the problem-solving process. *Psychological Review,* **45**: 107-133.

Lashley, K. S. (1929). *Brain Mechanisms and intelligence; a quantitative study of injuries to the brain.* Chicago: University of Chicago Press.

Lenneberg, E. H. (1967). *Biological Foundations of Language.* New York: Wiley.

Lewontin, R. C. (1981). On constraints and adaptation. *Behavoural and Brain Sciences,* **4**: 244-5.

Loeb, J. (1911). The Significance of Tropisms for Psychology. *Popular Science Monthly,* **79**: 105-25.

Lorenz, K (1971). [English translation; orig. 1950] Part and parcel in animal and human societies. In *Studies in animal and Human Behaviour*, Vol. II: 115-95. Cambridge, Mass: Harvard University Press.

Lorenz, K. (1965). *Evolution and Modification of behaviour.* Chicago: University of Chicago Press.

Macphail, E.M. (1982). *Brain and intelligence in Vertebrates.* Oxford: Clarendon Press.

McCulloch, T. L. & Pratt, J. C. (1934). A study of the pre-solutiom period in weight discrimination by white rats. *Journal of Comparative Psychology,* **18**: 271-290

McFarland, D. & Bösser, T. (1993).*Intelligent Behaviour in Animals and Robots.* Cambridge, Mass: MIT Press.

McGonigle, B.O. (1987). Non-verbal thinking by animals. *Nature* , **325**: 110-112

McGonigle, B. O. (1991). Incremental intelligent systems by design. In J. Arcady-Meyer and S. Wilson (Eds) *From Animals to animats.* Cambridge: MIT Press.

McGonigle, B.O. (1995). Review of L. Smith (1993) Necessary knowledge: Piagetian Perspectives on Constructivism. *British Journal of Developmental Psychology,*

McGonigle, B. O. & Chalmers, M. (1977). Are Monkeys Logical ? *Nature,* **267**: 694-696.

McGonigle, B. O. & Chalmers, M (1980). On the genesis of relational terms: A comparative study of monkeys and children. *Antropologica Contemporanea,* **3**: 236.

McGonigle, B. O. & Chalmers, M. (1984). The selective impact of question form and input mode on the symbolic distance effect in children. *Journal of Experimental Child Psychology,* **37**: 525-554.

McGonigle, B. O. & Chalmers, M. (1986). Representations and strategies during inference. In T. Myers, K. Brown & B. O. McGonigle (Eds) *Reasoning and Discourse Processes.* London: Academic Press.

McGonigle, B. O. & Chalmers, M. (1992). Monkeys are rational ! *The Quarterly Journal of Experimental Psychology,* 45B (3): 189-228

McGonigle, B. O. & Chalmers, M. (1993). An experimental analysis of ordering skills in children. ESRC Project grant final report. British Library, London.

McGonigle, B. O. & Chalmers, M. (1995). *The Growth of Intelligent Systems: A Cognitive Analysis.* Forthcoming: MIT Press.

McGonigle, B. O. & Chalmers, M. (1996). The Ontology of Order. In L. Smith (Ed.), *Critical readings on Piaget.* London: Routledge.

McGonigle, B. & Chalmers, M (1997a). Cognitive learning in monkeys and man. In S. Fountain (Ed), *Cognition in relation to the biomedical sciences.* Forthcoming: Ohio: Sage Publications

McGonigle, B. O. & Chalmers, M. (1997b). Rationality as Optimised Cognitive Self-regulation. In M. Oaksford & N. Chater (Eds), *Rational Models of Cognition.* Oxford: Oxford University Press.

McGonigle, B. O., de Lillo, C. & Dickinson, A. R. (1992). *Serial order induced search in children and monkeys.* Abstracts of the Vth European Conference on Developmental Psychology, Seville, Spain.

McGonigle, B. O., de Lillo, C. & Dickinson, A. R. (1994). *Classification to order: a comparative analysis of categorical seriation in monkey and man.* Abstracts of the XVth Congress of the International Primatological Society, Kuta, Bali, Indonesia.

McGonigle, B.O. & Jaswal, V. (1993). *Categorical Seriation in Children: A preliminary report.* Laboratory for Cognitive Neuroscience and Intelligent Systems, Edinburgh University. July, 1993.

McGonigle, B.O. & Jones, B.T. (1978). Levels of stimulus processing by the squirrel monkey: Relative and absolute judgements compared. *Perception,* **7**: 635-659.

Maier, N. R. F. & Schneirla, T. C. (1935). *Principles of Animal Psychology.* New York: McGraw-Hill.

Menzel, E. & Halperin, S.T. (1975). Purposive behaviour as a basis for objective communication between chimpanzees. *Science*, **189**: 652-4.

Miles, H.L. (1983) Apes and Language: the search for communicative competence. In J. de Luce & H.T. Wilder (Eds) *Language and Apes* pp. 43-61. New York: Springer-Verlag.

Miles, H.L. (1990) The cognitive foundations for reference in a signing orang utan. In S.T. Parker & K.R. Gibson (Eds) *"Language" and intelligence in monkeys and apes: Comparative developmental perspectives* pp. 511-539. New York: Cambridge University Press.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* **63**: 81-97.

Morgan, C.L. (1894). *An Introduction to Comparative Psychology.* London: Scott.

Morgan, C.L. (1896). *Habit and Instinct.* London: Edward Arnold.

Morgan, M.J., Fitch, M.D., Holman, J.D., & Lea, S.E. (1976). Pigeons learn the concept of 'A'. *Perception,* **5**: 57-66.

Newell, A. (1992). Precis of unified theories of cognition. *Behavioral and Brain Sciences*, **15**: 425-492

Paivio, A. (1978). A dual-coding approach to perception and cognition. In H. Pick and E. Saltzman (Eds), *Modes of Perceiving and Processing Information.* Hillside: Erlbaum.

Parker, S. T. (1990). Origins of comparative developmental evolutionary studies of primate mental abilities. In S. T. Parker and K. R. Gibson (Eds), *"Language" and intelligence in monkeys and apes: Comparative and developmental perspectives.* New York: Cambridge University Press.

Passingham, R. E. (1981). Primate specialisations in the brain and intelligence. *Symposium of the Zoological Society of London,* **46**: 361-388.

Passingham, R. E. (1982). *The Human Primate.* Oxford: W. H. Freeman.

Patterson, F.G. 1978). The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language* **5**: 72-97.

Pavlov, I. (1927). *Conditioned Reflexes.* (Trans. of G.V. Anrep) London: Oxford University

Pearce, J. M. (1987). *An Introduction to Animal Cognition.* Sussex: Lawrence Erlbaum.

Pepperberg, I.M. (1981). Functional vocalisations by an African grey parrot (*Psittacus erithacus*) . *Z. Tierpsychol.* **55**: 139-160.

Pepperberg, I.M. (1983). Cognition in the African grey parrot: preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning and Behaviour*, **11**: 179-185.

Piaget J. (1950). *The Psychology of Intelligence.* London: Routledge & Keegan Paul.

Piaget J. (1952). *The Origins of of Intelligence in the Child.* London: Routledge & Keegan Paul.

Piaget J. (1954). *The Contruction of Reality in the Child.* New York: Basic Books.

Piaget, J. (1971). *Biology and knowledge.* Edinburgh: Edinburgh University Press.

Piaget, J. & Inhelder, B. (1969). *The Psychology of the Child.* New York: Basic Books.

Plotkin, H. C. (1988). Behaviour and Evolution. In H. C. Plotkin (Ed.) *The role of Behaviour in Evolution.* Cambridge, Mass.: MIT Press.

Plotkin, H. C. & Olding-Smee, F. J. (1982). Learning in the context of a hierarchy of knowledge gaining processes. In H. C. Plotkin (Ed.) *Learning, Development and Culture: Essays in Evolutionary Epistemology.* Chichester: Wiley.

Potts, G. (1972). Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behaviour,* **11**: 727-740.

Povinelli, D.J. & Eddy, T.J. (1996). What chimpanzees know about seeing. *Monographs of the Society for Research in Child Development,* Vol. **61**, No. 2.

Premack, D. (1970). A functional analysis of language. *Journal of the Experimental Analysis of Behaviour,* **14**: 107-125.

Premack, D. (1971). Language in Chmpanzees ? *Science,* **172**: 808-822.

Premack, D. (1976). *Intelligence in Ape and Man.* Hillsdale, Lawrence-Erlbaum.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind ? *The Behavioural and Brain Sciences,* **4**: 515-526

Restle, F. (1958). Towards a quantitative description of learning set data. *Psychological Review,* **65**: 77-91.

Reese, H. W. (1968). *The Perception of Stimulus Relations: Discrimination learning and transposition.* New York: Academic Press

Romanes, G.J. (1882). *Animal Intelligence.* London: Kegan, Paul, Trench & Co.

Romanes, G.J. (1884). *Mental Evolution in Animals.* New York: Appleton.

Rumbaugh, D. M. (1977). *Language Learning by a Chimpanzee: The LANA project.* New York: Academic Press.

Russell, I. S. (1979). Brain size and Intelligence: a comparative perspective. In D. A. Oakley & H. C. Plotkin (Eds), *Brain, Behaviour and Evolution.*

Savage-Rumbaugh, E.S., Rumbaugh, D.M. & Boysen, S. (1978) Linguistically mediated tool use and exchange by chimpanzees. *The Behaviour and Brain Sciences,* **4**: 539-554.

Schaller, G.B. (1963) *The Mountain Gorilla: Ecology and Behaviour.* University of Chicago Press.

Schrier, A. M. (1966). Transfer by macaque monkeys between learning set and repeated reversal. *Perceptual and Motor Skills.* **23**, 787-92.

Schusterman, R.J. & Gisiner, R. (1984) Artificial language comprehension in dolphins and sea lions: The essential cognitive skills. *The Psychological Record.* **38**: 311-348.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review,* **77**: 406-418.

Seyfarth, R.M., Cheney, D.L. & Marler, P. (1980). Monkey responses to three different alarm calls: evidence for preditor classification and semantic communication. *Science*, **210**: 801-803.

Singer, P. (1990). *Animal liberation.* New York: Harper Collins.

Skinner, B. F. (1938). *The Behaviour of Organisms.* New York: Appleton-Century.

Skinner, B.F. (1969). *Contingencies of Reinforcement.* New York: Appleton-Century.

Skinner, B. F. (1957).*Verbal Behavior.* New York: Appleton-Century.

Smith, L. (1993). *Necessary Knowledge: Piagetian Perspectives on Constructivism.* Hove: Lawrence Erlbaum.

Spence, K. K. (1945). An experimental test on the continuity and non-continuity theories of discrimination learning. *Journal of Experimental Psychology,* **35**: 253-266.

Spencer, H. (1855). *Principles of Psychology* . 1st edn. London: Longman.

Stamp-Dawkins, H. (1992). *Through Our Eyes Only ?: the search for animal consciousness.* Oxford: Blackwell.

Tenaza, R. & Marler, P. (1977). Signalling behaviour of apes with special reference to vocalisation. In T.A. Seboek (Ed.) *How Animals Communicate,* pp 965-1033. Bloomington: Indiana University Press.

Terrace, H. S. (1986). A non-verbal organism's knowledge of ordinal position in a serial learning task. *Journal of Experimental Psychology: Animal Behaviour Processes*, **12**: 203-214.

Terrace, H. S. (1991). Chunking during serial learning by a pigeon: I. Basic Evidence. *Journal of Experimental Psychology: Animal Behaviour Processes*, **17**: 81-93.

Terrace, H. S. & Chen, S. (1991a). Chunking during serial learning by a pigeon: II Integrity of a chunk on a new list. *Journal of Experimental Psychology: Animal Behaviour Processes*, **17**: 94-106.

Terrace, H. S. & Chen, S. (1991b). Chunking during serial learning by a pigeon: III What are the necessary conditions for establishing a chunk ? *Journal of Experimental Psychology: Animal Behaviour Processes*, **17**: 107-118.

Terrace, H. S., Petitto, L. A., Sanders, R. J. & Bever, T. G. (1979). Can an Ape Create a Sentence ? *Science,* **200**: 891-902

Thorndyke, E.L. (1898). Animal Intelligence: an Experimental Study of the Associative Processes in Animals. *Psychological Review. Monograph Supplement* No. 28: 1-109.

Thorndyke, E.L. (1911). *Animal Intelligence.* New York: Macmillan.

Tinbergen, N. (1951). *The Study of Instinct.* Oxford: Clarendon Press.

Trabasso, T., Riley. C. A., & Wilson, E. G. (1975). The Representation of linear and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed) *Reasoning: Representation and processes in children and adults* (pp 201-229). Hillsdale, NJ: Lawrence Erlbaum.

Tuddenham, R. D. (1963). The Nature and measurement of intelligence. In L. Postman, (Ed), *Psychology in the Making.*

Waddington, C. H. (1969). Paradigm for an evolutionary process. In C. H. Waddington (Ed.) *Towards a theoretical biology,* Vol II: *Sketches.* Edinburgh: Edinburgh University Press.

Warren, J.M. (1965). Primate Learning in comparative Perspective. In A. M. Schrier, H. F. Harlow & F. Stollnitz (Eds), *Behaviour of non-human primates: modern research trends*, Vol 1: 249-81.

Watson, J.B. (1913). Psychology as the Behaviourist views it. *Psychological Review,* **20**, 158-77.

Watson, J.B. (1924). *Behaviourism.* New York: Norton.

Whiten, A. (1997) Personal communication.

Whiten, A. & Byrne, R.W. (1988). Tactical deception in primates. *The Behavioural and Brain sciences,* **11**: 233-44.

Whiten, A. & Byrne R.W. (1997) *Machiavellian intelligence II: Extentions and evaluations.* Cambridge University Press.

Woodruff, G. & Premack, D. (1979). Intentional communication in the chimpanzee. *Cognition*, **7**: 333-362.

Yerkes, R.M.(1916). The Mental Lives of Monkeys and Apes. *Behaviour Monographs*, **3**: No. 1. New York: Holt. (reprinted in Delmar, New York: Scholars Facsimiles and Reprints, 1979).

Yerkes, R.M. and Morgulis, S. (1909). The method of Pavlov in animal psychology. *Psychological Bulletin*, **6**: 257-273.